CORRECTED VERSION*

PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

# INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)


(51) International Patent Classification 7 : C12N 15/53, 9/02, 15/81, 5/10, C12Q 1/68, C12P 7/44, C12R 1/74 — A3

(11) International Publication Number: WO 00/20566

(43) International Publication Date: 13 April 2000 (13.04.00)

(21) International Application Number: PCT/US99/20797

(22) International Filing Date: 10 September 1999 (10.09.99)

(30) Priority Data:

| | | |
|---|---|---|
| 60/103,099 | 5 October 1998 (05.10.98) | US |
| 60/123,555 | 10 March 1999 (10.03.99) | US |



(71) Applicant *(for all designated States except US)*: HENKEL CORPORATION [US/US]; Suite 200, 2500 Renaissance Boulevard, Gulph Mills, PA 19406 (US).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: WILSON, C., Ron [–/US]; 6327 Belmont Road, Loveland, OH 45140 (US). CRAFT, David, L. [–/US]; 26 Rosewood Lane, Fort Thomas, KY 41075 (US). EIRICH, L., Dudley [–/US]; 227 W. Stoneridge Drive, Cincinnati, OH 45150 (US). ESHOO, Mark [–/US]; Apartment 5, 2821 Hillegass Avenue, Berkeley, CA 94705 (US). MADDURI, Krishna, M. [–/US]; Apartment C, 8237 Lenox Lane, Indianapolis, IN 46268 (US). CORNETT, Cathy, A. [–/US]; Apartment #11, 697 Meadow Wood Drive, Crescent Springs, KY 41017 (US). BRENNER, Alfred, A. [–/US]; 1563 Gamay Street, Santa Rosa, CA 95403 (US). TANG, Maria [–/US]; 2043 Cliffwood Drive, Fairfield, CA 94533 (US). LOPER, John, C. [–/US]; 6315 Parkman Place, Cincinnati, OH 45213 (US). GLEESON, Martin [–/US]; 4228 Cordobes Cove, San Diego, CA 92130 (US).

(74) Agents: DRACH, John, E. et al.; Henkel Corporation, Patent Dept., Suite 200, 2500 Renaissance Boulevard, Gulph Mills, PA 19406 (US).



(81) Designated States: AE, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZA, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SL, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

**Published**

*With international search report.*

(88) Date of publication of the international search report: 12 October 2000 (12.10.00)


(54) Title: CYTOCHROME P450 MONOOXYGENASE AND NADPH CYTOCHROME P450 OXIDOREDUCTASE GENES AND PROTEINS RELATED TO THE OMEGA HYDROXYLASE COMPLEX OF *CANDIDA TROPICALIS* AND METHODS RELATING THERETO

(57) Abstract


Novel genes have been isolated which encode cytochrome P450 and NADPH reductase enzyme of the $\omega$–hydroxylase complex of *C. tropicalis* 20336. Vectors including these genes, transfected host cells and transformed host cells are provided. Methods of producing of cytochrome P450 and NADPH reductase enzymes are also provided which involve transforming a host cell with a gene encoding these enzymes and culturing the cells. Methods of increasing the production of a dicarboxylic acid and methods of increasing production of the aforementioned enzymes are also provided which involve increasing in the host cell the number of genes encoding these enzymes. A method for discriminating members of a gene family by quantifying the expression of genes is also provided.



*(Referred to in PCT Gazette No. 34/2000, Section II)

PATENT COOPERATION TREATY

From the RECEIVING OFFICE

# PCT

To:

JOHN E. DRACH
HENKEL CORPORATION, PATENT DEPARTMENT
2500 RENAISSANCE BOULEVARD
SUITE 200
GULPH MILLS PA 19406

## NOTIFICATION OF THE INTERNATIONAL APPLICATION NUMBER AND OF THE INTERNATIONAL FILING DATE

(PCT Rule 20.5(c))

| | |
|---|---|
| Date of mailing (*day/month/year*) | 15 OCT 1999 |

**IMPORTANT NOTIFICATION**

| Applicant's or agent's file reference | | |
|---|---|---|
| M6308-3PCT ( | | |
| International application No. | International filing date (*day/month/year*) | Priority date (*day/month/year*) |
| PCT/US99/20797 | 10 SEP 99 | 05 OCT 98 |

Applicant HENKEL CORPORATION

Title of the invention CYTOCHROME P450 MONOOXYGENASE AND NADPH CYTOCHROME P450...

1. The applicant is hereby notified that the international application has been accorded the international application number and the international filing date indicated above.

2. The applicant is further notified that the record copy of the international application:

   ☑ was transmitted to the International Bureau on 15 OCT 1999.

   ☐ has not yet been transmitted to the International Bureau for the reason indicated below and a copy of this notification has been sent to the International Bureau*:

   ☐ because the necessary national security clearance has not yet been obtained.

   ☐ because (*reason to be specified*):

* The International Bureau monitors the transmittal of the record copy by the receiving Office and will notify the applicant (with Form PCT/IB/301) of its receipt. Should the record copy not have been received by the expiration of 14 months from the priority date, the International Bureau will notify the applicant (Rule 22.1(c)).

3. FOREIGN TRANSMITTAL LICENSE INFORMATION Completed by: [signature]

   ☐ Additional license for foreign transmittal not required. This subject matter is covered by a license already granted on the equivalent U.S. national application. Refer to that license for information concerning its scope.

   ☐ License for foreign transmittal not required. 37 CFR 5.11(e)(1) or 37 CFR 5.11(e)(2). However, a license may be required for additional subject matter. See 37 CFR 5.15(b).

   ☑ Foreign transmittal license granted. 35 U.S.C. 184; 37 CFR 5.11 on 9-23-99 (date):

   ☑ 37 CFR 5.15(a) ☐ 37 CFR 5.15(b)

| Name and mailing address of the receiving Office | Authorized officer |
|---|---|
| Assistant Commissioner for Patents<br>Box PCT<br>Washington, D.C. 20231 Attn: RO/US | Yolanda Harrod<br>PCT/Internat'l Appl Processing Div. |
| Facsimile No. | Telephone No. (703) 305-3670 |

Form PCT/RO/105 (July 1992)

# CYTOCHROME P450 MONOOXYGENASE AND NADPH CYTOCHROME P450 OXIDOREDUCTASE GENES AND PROTEINS RELATED TO THE OMEGA HYDROXYLASE COMPLEX OF *CANDIDA TROPICALIS* AND METHODS RELATING THERETO

## CROSS REFERENCE TO RELATED APPLICATIONS

This application claims priority to U.S. Provisional Application Serial No. 60/103,099 filed October 5, 1998, and U.S. Provisional Application Serial No. 60/083,798 filed May 1, 1998.

## BACKGROUND

### 1. Field of the Invention

The present invention relates to novel genes which encode enzymes of the ω-hydroxylase complex in yeast *Candida tropicalis* strains. In particular, the invention relates to novel genes encoding the cytochrome P450 and NADPH reductase enzymes of the ω-hydroxylase complex in yeast *Candida tropicalis*, and to a method of quantitating the expression of genes.

### 2. Description of the Related Art

Aliphatic dioic acids are versatile chemical intermediates useful as raw materials for the preparation of perfumes, polymers, adhesives and macrolid antibiotics. While several chemical routes to the synthesis of long-chain alpha, ω-dicarboxylic acids are available, the synthesis is not easy and most methods result in mixtures containing shorter chain lengths. As a result, extensive purification steps are necessary. While it is known that long-chain dioic acids can also be produced by microbial transformation of alkanes, fatty acids or esters thereof, chemical synthesis has remained the most commercially viable route, due to limitations with the current biological approaches.

Several strains of yeast are known to excrete alpha, ω-dicarboxylic acids as a byproduct when cultured on alkanes or fatty acids as the carbon source. In particular, yeast belonging to the Genus *Candida*, such as *C. albicans, C. cloacae, C. guillermondii, C. intermedia, C. lipolytica, C. maltosa, C. parapsilosis* and *C. zeylenoides* are known to produce

such dicarboxylic acids (*Agr. Biol. Chem.* 35: 2033-2042 (1971)). Also, various strains of *C. tropicalis* are known to produce dicarboxylic acids ranging in chain lengths from $C_{11}$ through $C_{18}$ (Okino et al., BM Lawrence, BD Mookherjee and BJ Willis (eds), in *Flavors and Fragrances: A World Perspective*. Proceedings of the 10th International Conference of Essential Oils, Flavors and Fragrances, Elsevier Science Publishers BV Amsterdam (1988)), and are the basis of several patents as reviewed by Bühler and Schindler, in *Aliphatic Hydrocarbons in Biotechnology*, H. J. Rehm and G. Reed (eds), Vol. 169, Verlag Chemie, Weinheim (1984).

Studies of the biochemical processes by which yeasts metabolize alkanes and fatty acids have revealed three types of oxidation reactions: α-oxidation of alkanes to alcohols, ω-oxidation of fatty acids to alpha, ω-dicarboxylic acids and the degradative *β*-oxidation of fatty acids to $CO_2$ and water. The first two types of oxidations are catalyzed by microsomal enzymes while the last type takes place in the peroxisomes. In *C. tropicalis,* the first step in the ω-oxidation pathway is catalyzed by a membrane-bound enzyme complex (ω-hydroxylase complex) including a cytochrome P450 monooxygenase and a NADPH cytochrome reductase. This hydroxylase complex is responsible for the primary oxidation of the terminal methyl group in alkanes and fatty acids (Gilewicz et al., *Can. J. Microbiol.* 25:201 (1979)). The genes which encode the cytochrome P450 and NADPH reductase components of the complex have previously been identified as P450ALK and P450RED respectively, and have also been cloned and sequenced (Sanglard et al., *Gene* 76:121-136 (1989)). P450ALK has also been designated P450ALK1. More recently, ALK genes have been designated by the symbol *CYP* and RED genes have been designated by the symbol *CPR*. See, e.g., Nelson, *Pharmacogenetics* 6(1):1-42 (1996), which is incorporated herein by reference. See also Ohkuma et al., *DNA and Cell Biology* 14:163-173 (1995), Seghezzi et al., *DNA and Cell Biology*, 11:767-780 (1992) and Kargel et al., *Yeast* 12:333-348 (1996), each incorporated herein by reference. For example, P450ALK is also designated *CYP52* according to the nomenclature of Nelson, *supra*. Fatty acids are ultimately formed from alkanes after two additional oxidation steps, catalyzed by alcohol oxidase (Kemp et al., *Appl. Microbiol. and Biotechnol.* 28: 370-374 (1988)) and aldehyde dehydrogenase. The fatty acids can be further oxidized through the same or similar pathway to the corresponding dicarboxylic acid. The ω-oxidation of fatty acids proceeds via the ω-hydroxy fatty acid and its aldehyde derivative, to the corresponding dicarboxylic acid without the requirement for CoA activation. However, both fatty acids and dicarboxylic acids can be

degraded, after activation to the corresponding acyl-CoA ester through the β-oxidation pathway in the peroxisomes, leading to chain shortening. In mammalian systems, both fatty acid and dicarboxylic acid products of ω-oxidation are activated to their CoA-esters at equal rates and are substrates for both mitochondrial and peroxisomal β-oxidation (*J. Biochem.*, 102:225-234 (1987)). In yeast, β-oxidation takes place solely in the peroxisomes (*Agr.Biol.Chem.* 49:1821-1828 (1985)).

The production of dicarboxylic acids by fermentation of unsaturated $C_{14}$-$C_{16}$ monocarboxylic acids using a strain of the species *C. tropicalis* is disclosed in U.S. Patent 4,474,882. The unsaturated dicarboxylic acids correspond to the starting materials in the number and position of the double bonds. Similar processes in which other special microorganisms are used are described in U.S. Patents 3,975,234 and 4,339,536, in British Patent Specification 1,405,026 and in German Patent Publications 21 64 626, 28 53 847, 29 37 292, 29 51 177, and 21 40 133.

Cytochromes P450 (P450s) are terminal monooxidases of a multicomponent enzyme system as described above. They comprise a superfamily of proteins which exist widely in nature having been isolated from a variety of organisms as described e.g., in Nelson, *supra*. These organisms include various mammals, fish, invertebrates, plants, mollusk, crustaceans, lower eukaryotes and bacteria (Nelson, *supra)*. First discovered in rodent liver microsomes as a carbon-monoxide binding pigment as described, e.g., in Garfinkel, *Arch. Biochem. Biophys*. 77:493-509 (1958), which is incorporated herein by reference, P450s were later named based on their absorption at 450 nm in a reduced-CO coupled difference spectrum as described, e.g., in Omura et al., *J. Biol. Chem*. 239:2370-2378 (1964), which is incorporated herein by reference.

P450s catalyze the metabolism of a variety of endogenous and exogenous compounds (Nelson, *supra*). Endogenous compounds include steroids, prostanoids, eicosanoids, fat-soluble vitamins, fatty acids, mammalian alkaloids, leukotrines, biogenic amines and phytolexins (Nelson, *supra*). P450 metabolism involves such reactions as epoxidation, hydroxylation, deakylation, N-hydroxylation, sulfoxidation, desulfuration and reductive dehalogenation. These reactions generally make the compound more water soluble, which is conducive for excretion, and more electrophilic. These electrophilic products can have detrimental effects if they react with DNA or other cellular constituents. However, they can react

through conjugation with low molecular weight hydrophilic substances resulting in glucoronidation, sulfation, acetylation, amino acid conjugation or glutathione conjugation typically leading to inactivation and elimination as described, e.g., in Klaassen et al., *Toxicology*, 3rd ed, Macmillan, New York, 1986, incorporated herein by reference.

P450s are heme thiolate proteins consisting of a heme moiety bound to a single polypeptide chain of 45,000 to 55,000 Da. The iron of the heme prosthetic group is located at the center of a protoporphyrin ring. Four ligands of the heme iron can be attributed to the porphyrin ring. The fifth ligand is a thiolate anion from a cysteinyl residue of the polypeptide. The sixth ligand is probably a hydroxyl group from an amino acid residue, or a moiety with a similar field strength such as a water molecule as described, e.g., in Goeptar et al., *Critical Reviews in Toxicology* 25(1):25-65 (1995), incorporated herein by reference.

Monooxygenation reactions catalyzed by cytochromes P450 in a eukaryotic membrane-bound system require the transfer of electrons from NADPH to P450 via NADPH-cytochrome P450 reductase (*CPR*) as described, e.g., in Taniguchi et al., *Arch. Biochem. Biophys.* 232:585 (1984), incorporated herein by reference. *CPR* is a flavoprotein of approximately 78,000 Da containing 1 mol of flavin adenine dinucleotide (FAD) and 1 mol of flavin mononucleotide (FMN) per mole of enzyme as described, e.g., in Potter et al., *J. Biol. Chem.* 258:6906 (1983), incorporated herein by reference. The FAD moiety of *CPR* is the site of electron entry into the enzyme, whereas FMN is the electron-donating site to P450 as described, e.g., in Vermilion et al., *J. Biol. Chem.* 253:8812 (1978), incorporated herein by reference. The overall reaction is as follows:

$$H^+ + RH + NADPH + O_2 \rightarrow ROH + NADP^+ + H_2O$$

Binding of a substrate to the catalytic site of P450 apparently results in a conformational change initiating electron transfer from *CPR* to P450. Subsequent to the transfer of the first electron, $O_2$ binds to the $Fe_2^+$-P450 substrate complex to form $Fe_3^+$ -P450-substrate complex. This complex is then reduced by a second electron from *CPR*, or, in some cases, NADH via cytochrome b5 and NADH-cytochrome b5 reductase as described, e.g., in Guengerich et al., *Arch. Biochem. Biophys.* 205:365 (1980), incorporated herein by reference. One atom of this reactive oxygen is introduced into the substrate, while the other is reduced to water. The

oxygenated substrate then dissociates, regenerating the oxidized form of the cytochrome P450 as described, e.g., in Klassen, Amdur and Doull, *Casarett and Doull's Toxicology*, Macmillan, New York (1986), incorporated herein by reference.

The P450 reaction cycle can be short-circuited in such a way that $O_2$ is reduced to $O_2^-$ and/or $H_2O_2$ instead of being utilized for substrate oxygenation. This side reaction is often referred to as the "uncoupling" of cytochrome P450 as described, e.g., in Kuthen et al., *Eur. J. Biochem.* 126:583 (1982) and Poulos et al., *FASEB* J. 6:674 (1992), both of which are incorporated herein by reference. The formation of these oxygen radicals may lead to oxidative cell damage as described, e.g., in Mukhopadhyay, *J. Biol. Chem.* 269(18):13390-13397 (1994) and Ross et al., *Biochem. Pharm.* 49(7):979-989 (1995), both of which are incorporated herein by reference. It has been proposed that cytochrome b5's effect on P450 binding to the *CPR* results in a more stable complex which is less likely to become "uncoupled" as described, e.g., in Yamazaki et al., *Arch. Biochem. Biophys.* 325(2):174-182 (1996), incorporated herein by reference.

P450 families are assigned based upon protein sequence comparisons. Notwithstanding a certain amount of heterogeneity, a practical classification of P450s into families can be obtained based on deduced amino acid sequence similarity. P450s with amino acid sequence similarity of between about 40 - 80% are considered to be in the same family, with sequences of about > 55% belonging to the same subfamily. Those with sequence similarity of about < 40% are generally listed as members of different P450 gene families (Nelson, *supra*). A value of about > 97% is taken to indicate allelic variants of the same gene, unless proven otherwise based on catalytic activity, sequence divergence in non-translated regions of the gene sequence, or chromosomal mapping.

The most highly conserved region is the HR2 consensus containing the invariant cysteine residue near the carboxyl terminus which is required for heme binding as described, e.g., in Gotoh et al. *J. Biochem.* 93:807-817 (1983) and Motohashi et al., *J. Biochem.* 101:879-997 (1987), both of which are incorporated herein by reference. Additional consensus regions, including the central region of helix I and the transmembrane region, have also been identified, as described, e.g, in Goeptar et al., *supra* and Kalb et al., *PNAS.* 85:7221-7225 (1988), incorporated herein by reference, although the HR2 cysteine is the only invariant amino acid among P450s.

Short chain (≤C12) aliphatic dicarboxylic acids (diacids) are important industrial intermediates in the manufacture of diesters and polymers, and find application as thermoplastics, plasticizing agents, lubricants, hydraulic fluids, agricultural chemicals, pharmaceuticals, dyes, surfactants, and adhesives. The high price and limited availability of short chain diacids are due to constraints imposed by the existing chemical synthesis.

Long-chain diacids (aliphatic α, ω-dicarboxylic acids with carbon numbers of 12 or greater, hereafter also referred to as diacids) ($HOOC-(CH_2)_n-COOH$) are a versatile family of chemicals with demonstrated and potential utility in a variety of chemical products including plastics, adhesives, and fragrances. Unfortunately, the full market potential of diacids has not been realized because chemical processes produce only a limited range of these materials at a relatively high price. In addition, chemical processes for the production of diacids have a number of limitations and disadvantages. All the chemical processes are restricted to the production of diacids of specific carbon chain lengths. For example, the dodecanedioic acid process starts with butadiene. The resulting product diacids are limited to multiples of four-carbon lengths and, in practice, only dodecanedioic acid is made. The dodecanedioic process is based on nonrenewable petrochemical feedstocks. The multireaction conversion process produces unwanted byproducts, which result in yield losses, $NO_X$ pollution and heavy metal wastes.

Long-chain diacids offer potential advantages over shorter chain diacids, but their high selling price and limited commercial availability prevent widespread growth in many of these applications. Biocatalysis offers an innovative way to overcome these limitations with a process that produces a wide range of diacid products from renewable feedstocks. However, there is no commercially viable bioprocess to produce long chain diacids from renewable resources.

## SUMMARY OF THE INVENTION

An isolated nucleic acid is provided which encodes a *CPRA* protein having the amino acid sequence set forth in SEQ ID NO: 83. An isolated nucleic acid is also provided which includes a coding region defined by nucleotides 1006-3042 as set forth in SEQ ID NO: 81. An isolated protein is provided which includes an amino acid sequence as set forth in SEQ ID NO: 83. A vector is provided which includes a nucleotide sequence encoding *CPRA* protein

including an amino acid sequence as set forth in SEQ ID NO: 83. A host cell is provided which is transfected or transformed with the nucleic acid encoding *CPRA* protein having an amino acid sequence as set forth in SEQ ID NO: 83. A method of producing a *CPRA* protein including an amino acid sequence as set forth in SEQ ID NO: 83 is also provided which includes a) transforming a suitable host cell with a DNA sequence that encodes the protein having the amino acid sequence as set forth in SEQ ID NO: 83; and b) culturing the cell under conditions favoring the expression of the protein.

An isolated nucleic acid is provided which encodes a *CPRB* protein having the amino acid sequence set forth in SEQ ID NO: 84. An isolated nucleic acid is provided which includes a coding region defined by nucleotides 1033-3069 as set forth in SEQ ID NO: 82. An isolated protein is provided which includes an amino acid sequence as set forth in SEQ ID NO: 84. A vector is provided which includes a nucleotide sequence encoding *CPRB* protein including an amino acid sequence as set forth in SEQ ID NO: 84. A host cell is provided which is transfected or transformed with the nucleic acid encoding *CPRB* protein having an amino acid sequence as set forth in SEQ ID NO: 84. A method of producing a *CPRB* protein including an amino acid sequence as set forth in SEQ ID NO: 84 is provided which includes a) transforming a suitable host cell with a DNA sequence that encodes the protein having the amino acid sequence as set forth in SEQ ID NO: 84; and b) culturing the cell under conditions favoring the expression of the protein.

An isolated nucleic acid is provided which encodes a *CYP52A1A* protein having the amino acid sequence set forth in SEQ ID NO: 95. An isolated nucleic acid is provided which includes a coding region defined by nucleotides 1177-2748 as set forth in SEQ ID NO: 85. An isolated protein is provided which includes an amino acid sequence as set forth in SEQ ID NO: 95. A vector is provided which includes a nucleotide sequence encoding *CYP52A1A* protein including an amino acid sequence as set forth in SEQ ID NO: 95. A host cell is provided which is transfected or transformed with the nucleic acid encoding *CYP52A1A* protein having an amino acid sequence as set forth in SEQ ID NO: 95. A method of producing a *CYP52A1A* protein including an amino acid sequence as set forth in SEQ ID NO: 95 is provided which includes a) transforming a suitable host cell with a DNA sequence that encodes the protein having the amino acid sequence as set forth in SEQ ID NO: 95; and b) culturing the cell under conditions favoring the expression of the protein.

An isolated nucleic acid encoding a *CYP52A2A* protein is provided which has the amino acid sequence set forth in SEQ ID NO: 96. An isolated nucleic acid is provided which includes a coding region defined by nucleotides 1199-2767 as set forth in SEQ ID NO: 86. An isolated protein is provided which includes an amino acid sequence as set forth in SEQ ID NO: 96. A vector is provided which includes a nucleotide sequence encoding *CYP52A2A* protein including an amino acid sequence as set forth in SEQ ID NO: 96. A host cell is provided which is transfected or transformed with the nucleic acid encoding *CYP52A2A* protein having an amino acid sequence as set forth in SEQ ID NO: 96. A method of producing a *CYP52A2A* protein including an amino acid sequence as set forth in SEQ ID NO: 96 is provided which includes a) transforming a suitable host cell with a DNA sequence that encodes the protein having the amino acid sequence as set forth in SEQ ID NO: 96; and b) culturing the cell under conditions favoring the expression of the protein.

An isolated nucleic acid encoding a *CYP52A2B* protein is provided which has the amino acid sequence set forth in SEQ ID NO: 97. An isolated nucleic acid is provided which includes a coding region defined by nucleotides 1072-2640 as set forth in SEQ ID NO: 87. An isolated protein is provided which includes an amino acid sequence as set forth in SEQ ID NO: 97. A vector is provided which includes a nucleotide sequence encoding *CYP52A2B* protein including an amino acid sequence as set forth in SEQ ID NO: 97. A host cell is provided which is transfected or transformed with the nucleic acid encoding *CYP52A2B* protein having an amino acid sequence as set forth in SEQ ID NO: 97. A method of producing a *CYP52A2B* protein including an amino acid sequence as set forth in SEQ ID NO: 97 is provided which includes a) transforming a suitable host cell with a DNA sequence that encodes the protein having the amino acid sequence as set forth in SEQ ID NO: 97; and b) culturing the cell under conditions favoring the expression of the protein.

An isolated nucleic acid encoding a *CYP52A3A* protein is provided which has the amino acid sequence set forth in SEQ ID NO: 98. An isolated nucleic acid is provided which includes a coding region defined by nucleotides 1126-2748 as set forth in SEQ ID NO: 88. An isolated protein is provided which includes an amino acid sequence as set forth in SEQ ID NO: 98. A vector is provided which includes a nucleotide sequence encoding *CYP52A3A* protein including an amino acid sequence as set forth in SEQ ID NO: 98. A host cell is provided which is transfected or transformed with the nucleic acid encoding *CYP52A3A* protein having an

amino acid sequence as set forth in SEQ ID NO: 98. A method of producing a *CYP52A3A* protein including an amino acid sequence as set forth in SEQ ID NO: 98 is provided which includes a) transforming a suitable host cell with a DNA sequence that encodes the protein having the amino acid sequence as set forth in SEQ ID NO: 98; and b) culturing the cell under conditions favoring the expression of the protein.

An isolated nucleic acid encoding a *CYP52A3B* protein is provided having the amino acid sequence as set forth in SEQ ID NO: 99. An isolated nucleic acid is provided which includes a coding region defined by nucleotides 913-2535 as set forth in SEQ ID NO: 89. An isolated protein is provided which includes an amino acid sequence as set forth in SEQ ID NO: 99. A vector is provided which includes a nucleotide sequence encoding *CYP52A3B* protein including an amino acid sequence as set forth in SEQ ID NO: 99. A host cell is provided which is transfected or transformed with the nucleic acid encoding *CYP52A3B* protein having an amino acid sequence as set forth in SEQ ID NO: 99. A method of producing a *CYP52A3B* protein including an amino acid sequence as set forth in SEQ ID NO: 99 is provided which includes a) transforming a suitable host cell with a DNA sequence that encodes the protein having the amino acid sequence as set forth in SEQ ID NO: 99; and b) culturing the cell under conditions favoring the expression of the protein.

An isolated nucleic acid encoding a *CYP52A5A* protein is provided having the amino acid sequence set forth in SEQ ID NO: 100. An isolated nucleic acid is provided which includes a coding region defined by nucleotides 1103-2656 as set forth in SEQ ID NO: 90. An isolated protein is provided which includes an amino acid sequence as set forth in SEQ ID NO: 100. A vector is provided which includes a nucleotide sequence encoding *CYP52A5A* protein including an amino acid sequence as set forth in SEQ ID NO: 100. A host cell is provided which is transfected or transformed with the nucleic acid encoding *CYP52A5A* protein having an amino acid sequence as set forth in SEQ ID NO: 100. A method of producing a *CYP52A5A* protein including an amino acid sequence as set forth in SEQ ID NO: 100 is provided which includes a) transforming a suitable host cell with a DNA sequence that encodes the protein having the amino acid sequence as set forth in SEQ ID NO: 100; and b) culturing the cell under conditions favoring the expression of the protein.

An isolated nucleic acid encoding a *CYP52A5B* protein is provided having the amino acid sequence as set forth in SEQ ID NO: 101. An isolated nucleic acid is provided

which includes a coding region defined by nucleotides 1142-2695 as set forth in SEQ ID NO: 91. An isolated protein is provided which includes an amino acid sequence as set forth in SEQ ID NO: 101. A vector is provided which includes a nucleotide sequence encoding *CYP52A5B* protein including the amino acid sequence as set forth in SEQ ID NO: 101. A host cell is provided which is transfected or transformed with the nucleic acid encoding *CYP52A5B* protein having the amino acid sequence as set forth in SEQ ID NO: 101. A method of producing a *CYP52A5B* protein including an amino acid sequence as set forth in SEQ ID NO: 101 is provided which includes a) transforming a suitable host cell with a DNA sequence that encodes the protein having the amino acid sequence as set forth in SEQ ID NO: 101; and b) culturing the cell under conditions favoring the expression of the protein.

An isolated nucleic acid encoding a *CYP52A8A* protein is provided having the amino acid sequence set forth in SEQ ID NO: 102. An isolated nucleic acid is provided which includes a coding region defined by nucleotides 464-2002 as set forth in SEQ ID NO: 92. An isolated protein is provided which includes an amino acid sequence as set forth in SEQ ID NO: 102. A vector is provided which includes a nucleotide sequence encoding *CYP52A8A* protein including an amino acid sequence as set forth in SEQ ID NO: 102. A host cell is provided which is transfected or transformed with the nucleic acid encoding *CYP52A8A* protein having an amino acid sequence as set forth in SEQ ID NO: 102. A method of producing a *CYP52A8A* protein including an amino acid sequence as set forth in SEQ ID NO: 102 is provided which includes a) transforming a suitable host cell with a DNA sequence that encodes the protein having the amino acid sequence as set forth in SEQ ID NO: 102; and b) culturing the cell under conditions favoring the expression of the protein.

An isolated nucleic acid encoding a *CYP52A8B* protein is provided having the amino acid sequence set forth in SEQ ID NO: 103. An isolated nucleic acid is provided which includes a coding region defined by nucleotides 1017-2555 as set forth in SEQ ID NO: 93. An isolated protein is provided which includes an amino acid sequence as set forth in SEQ ID NO: 103. A vector is provided which includes a nucleotide sequence encoding *CYP52A8B* protein including an amino acid sequence as set forth in SEQ ID NO: 103. A host cell is provided which is transfected or transformed with the nucleic acid encoding *CYP52A8B* protein having an amino acid sequence as set forth in SEQ ID NO: 103. A method of producing a *CYP52A8B* protein including an amino acid sequence as set forth in SEQ ID NO: 103 is provided which

includes a) transforming a suitable host cell with a DNA sequence that encodes the protein having the amino acid sequence as set forth in SEQ ID NO: 103; and b) culturing the cell under conditions favoring the expression of the protein.

An isolated nucleic acid encoding a *CYP52D4A* protein is provided having the amino acid sequence set forth in SEQ ID NO: 104. An isolated nucleic acid is provided including a coding region defined by nucleotides 767-2266 as set forth in SEQ ID NO: 94. An isolated protein is provided which includes an amino acid sequence as set forth in SEQ ID NO: 104. A vector is provided which includes a nucleotide sequence encoding *CYP52D4A* protein including an amino acid sequence as set forth in SEQ ID NO: 104. A host cell is provided which is transfected or transformed with the nucleic acid encoding *CYP52D4A* protein having an amino acid sequence as set forth in SEQ ID NO: 104. A method of producing a *CYP52D4A* protein including an amino acid sequence as set forth in SEQ ID NO: 104 is provided which includes a) transforming a suitable host cell with a DNA sequence that encodes the protein having the amino acid sequence as set forth in SEQ ID NO: 104; and b) culturing the cell under conditions favoring the expression of the protein.

A method for discriminating members of a gene family by quantifying the amount of target mRNA in a sample is provided which includes a) providing an organism containing a target gene; b) culturing the organism with an organic substrate which causes upregulation in the activity of the target gene; c) obtaining a sample of total RNA from the organism at a first point in time; d) combining at least a portion of the sample of the total RNA with a known amount of competitor RNA to form an RNA mixture, wherein the competitor RNA is substantially similar to the target mRNA but has a lesser number of nucleotides compared to the target mRNA; e) adding reverse transcriptase to the RNA mixture in a quantity sufficient to form corresponding target DNA and competitor DNA; (f) conducting a polymerase chain reaction in the presence of at least one primer specific for at least one substantially non-homologous region of the target DNA within the gene family, the primer also specific for the competitor DNA; g) repeating steps (c-f) using increasing amounts of the competitor RNA while maintaining a substantially constant amount of target RNA; h) determining the point at which the amount of target DNA is substantially equal to the amount of competitor DNA; i) quantifying the results by comparing the ratio of the concentration of unknown target to the known concentration of competitor; and j)

obtaining a sample of total RNA from the organism at another point in time and repeating steps (d-i).

A method for increasing production of a dicarboxylic acid is provided which includes a) providing a host cell having a naturally occurring number of *CPRA* genes; b) increasing, in the host cell, the number of *CPRA* genes which encode a *CPRA* protein having the amino acid sequence as set forth in SEQ ID NO: 83; c) culturing the host cell in media containing an organic substrate which upregulates the *CPRA* gene, to effect increased production of dicarboxylic acid.

A method for increasing the production of a *CPRA* protein having an amino acid sequence as set forth in SEQ ID NO: 83 is provided which includes a) transforming a host cell having a naturally occurring amount of *CPRA* protein with an increased copy number of a *CPRA* gene that encodes the *CPRA* protein having the amino acid sequence as set forth in SEQ ID NO: 83; and b) culturing the cell and thereby increasing expression of the protein compared with that of a host cell containing a naturally occurring copy number of the *CPRA* gene.

A method for increasing production of a dicarboxylic acid is provided which includes a) providing a host cell having a naturally occurring number of *CPRB* genes; b) increasing, in the host cell, the number of *CPRB* genes which encode a *CPRB* protein having the amino acid sequence as set forth in SEQ ID NO: 84; c) culturing the host cell in media containing an organic substrate which upregulates the *CPRB* gene, to effect increased production of dicarboxylic acid.

A method for increasing the production of a *CPRB* protein having an amino acid sequence as set forth in SEQ ID NO: 84 is provided which includes a) transforming a host cell having a naturally occurring amount of *CPRB* protein with an increased copy number of a *CPRB* gene that encodes the *CPRB* protein having the amino acid sequence as set forth in SEQ ID NO: 84; and b) culturing the cell and thereby increasing expression of the protein compared with that of a host cell containing a naturally occurring copy number of the *CPRB* gene.

A method for increasing production of a dicarboxylic acid is provided which includes a) providing a host cell having a naturally occurring number of *CYP52A1A* genes; b) increasing, in the host cell, the number of *CYP52A1A* genes which encode a *CYP52A1A* protein having the amino acid sequence as set forth in SEQ ID NO: 95; c) culturing the host cell in media

containing an organic substrate which upregulates the *CYP52A1A* gene, to effect increased production of dicarboxylic acid.

A method for increasing the production of a *CYP52A1A* protein having an amino acid sequence as set forth in SEQ ID NO: 95 is provided which includes a) transforming a host cell having a naturally occurring amount of *CYP52A1A* protein with an increased copy number of a *CYP52A1A* gene that encodes the *CYP52A1A* protein having the amino acid sequence as set forth in SEQ ID NO: 95; and b) culturing the cell and thereby increasing expression of the protein compared with that of a host cell containing a naturally occurring copy number of the *CYP52A1A* gene.

A method for increasing production of a dicarboxylic acid is provided which includes a) providing a host cell having a naturally occurring number of *CYP52A2A* genes; b) increasing, in the host cell, the number of *CYP52A2A* genes which encode a *CYP52A2A* protein having the amino acid sequence as set forth in SEQ ID NO: 96; c) culturing the host cell in media containing an organic substrate which upregulates the *CYP52A2A* gene, to effect increased production of dicarboxylic acid.

A method for increasing the production of a *CYP52A2A* protein having an amino acid sequence as set forth in SEQ ID NO: 96 is provided which includes a) transforming a host cell having a naturally occurring amount of *CYP52A2A* protein with an increased copy number of a *CYP52A2A* gene that encodes the *CYP52A2A* protein having the amino acid sequence as set forth in SEQ ID NO: 96; and b) culturing the cell and thereby increasing expression of the protein compared with that of a host cell containing a naturally occurring copy number of the *CYP52A2A* gene.

A method for increasing production of a dicarboxylic acid is provided which includes a) providing a host cell having a naturally occurring number of *CYP52A2B* genes; b) increasing, in the host cell, the number of *CYP52A2B* genes which encode a *CYP52A2B* protein having the amino acid sequence as set forth in SEQ ID NO: 97; c) culturing the host cell in media containing an organic substrate which upregulates the *CYP52A2B* gene, to effect increased production of dicarboxylic acid.

A method for increasing the production of a *CYP52A2B* protein having an amino acid sequence as set forth in SEQ ID NO: 97 is provided which includes a) transforming a host cell having a naturally occurring amount of *CYP52A2B* protein with an increased copy number of

a *CYP52A2B* gene that encodes the *CYP52A2B* protein having the amino acid sequence as set forth in SEQ ID NO: 97; and b) culturing the cell and thereby increasing expression of the protein compared with that of a host cell containing a naturally occurring copy number of the *CYP52A2B* gene.

A method for increasing production of a dicarboxylic acid is provided which includes a) providing a host cell having a naturally occurring number of *CYP52A3A* genes; b) increasing, in the host cell, the number of *CYP52A3A* genes which encode a *CYP52A3A* protein having the amino acid sequence as set forth in SEQ ID NO: 98; c) culturing the host cell in media containing an organic substrate which upregulates *CYP52A3A* gene, to effect increased production of dicarboxylic acid.

A method for increasing the production of a *CYP52A3A* protein having an amino acid sequence as set forth in SEQ ID NO: 98 is provided which includes a) transforming a host cell having a naturally occurring amount of *CYP52A3A* protein with an increased copy number of a *CYP52A3A* gene that encodes the *CYP52A3A* protein having the amino acid sequence as set forth in SEQ ID NO: 98; and b) culturing the cell and thereby increasing expression of the protein compared with that of a host cell containing a naturally occurring copy number of the *CYP52A3A* gene.

A method for increasing production of a dicarboxylic acid is provided which includes a) providing a host cell having a naturally occurring number of *CYP52A3B* genes; b) increasing, in the host cell, the number of *CYP52A3B* genes which encode a *CYP52A3B* protein having the amino acid sequence as set forth in SEQ ID NO: 99; c) culturing the host cell in media containing an organic substrate which upregulates the *CYP52A3B* gene, to effect increased production of dicarboxylic acid.

A method for increasing the production of a *CYP52A3B* protein having an amino acid sequence as set forth in SEQ ID NO: 99 is provided which includes a) transforming a host cell having a naturally occurring amount of *CYP52A3B* protein with an increased copy number of a *CYP52A3B* gene that encodes the *CYP52A3B* protein having the amino acid sequence as set forth in SEQ ID NO: 99; and b) culturing the cell and thereby increasing expression of the protein compared with that of a host cell containing a naturally occurring copy number of the *CYP52A3B* gene.

A method for increasing production of a dicarboxylic acid is provided which includes a) providing a host cell having a naturally occurring number of *CYP52A5A* genes; b) increasing, in the host cell, the number of *CYP52A5A* genes which encode a *CYP52A5A* protein having the amino acid sequence as set forth in SEQ ID NO: 100; c) culturing the host cell in media containing an organic substrate which upregulates the *CYP52A5A* gene, to effect increased production of dicarboxylic acid.

A method for increasing the production of a *CYP52A5A* protein having an amino acid sequence as set forth in SEQ ID NO: 100 is provided which includes a) transforming a host cell having a naturally occurring amount of *CYP52A5A* protein with an increased copy number of a *CYP52A5A* gene that encodes the *CYP52A5A* protein having the amino acid sequence as set forth in SEQ ID NO: 100; and b) culturing the cell and thereby increasing expression of the protein compared with that of a host cell containing a naturally occurring copy number of the *CYP52A5A* gene.

A method for increasing production of a dicarboxylic acid is provided which includes a) providing a host cell having a naturally occurring number of *CYP52A5B* genes; b) increasing, in the host cell, the number of *CYP52A5B* genes which encode a *CYP52A5B* protein having the amino acid sequence as set forth in SEQ ID NO: 101; c) culturing the host cell in media containing an organic substrate which upregulates the *CYP52A5B* gene, to effect increased production of dicarboxylic acid.

A method for increasing the production of a *CYP52A5B* protein having an amino acid sequence as set forth in SEQ ID NO: 101 is provided which includes a) transforming a host cell having a naturally occurring amount of *CYP52A5B* protein with an increased copy number of a *CYP52A5B* gene that encodes the *CYP52A5B* protein having the amino acid sequence as set forth in SEQ ID NO: 101; and b) culturing the cell and thereby increasing expression of the protein compared with that of a host cell containing a naturally occurring copy number of the *CYP52A5B* gene.

A method for increasing production of a dicarboxylic acid is provided which includes a) providing a host cell having a naturally occurring number of *CYP52A8A* genes; b) increasing, in the host cell, the number of *CYP52A8A* genes which encode a *CYP52A8A* protein having the amino acid sequence as set forth in SEQ ID NO: 102; c) culturing the host cell in

media containing an organic substrate which upregulates the *CYP52A8A* gene, to effect increased production of dicarboxylic acid.

A method for increasing the production of a *CYP52A8A* protein having an amino acid sequence as set forth in SEQ ID NO: 102 is provided which includes a) transforming a host cell having a naturally occurring amount of *CYP52A8A* protein with an increased copy number of a *CYP52A8A* gene that encodes the *CYP52A8A* protein having the amino acid sequence as set forth in SEQ ID NO: 102; and b) culturing the cell and thereby increasing expression of the protein compared with that of a host cell containing a naturally occurring copy number of the *CYP52A8A* gene.

A method for increasing production of a dicarboxylic acid is provided which includes a) providing a host cell having a naturally occurring number of *CYP52A8B* genes; b) increasing, in the host cell, the number of *CYP52A8B* genes which encode a *CYP52A8B* protein having the amino acid sequence as set forth in SEQ ID NO: 103; c) culturing the host cell in media containing an organic substrate which upregulates the *CYP52A8B* gene, to effect increased production of dicarboxylic acid.

A method for increasing the production of a *CYP52A8B* protein having an amino acid sequence as set forth in SEQ ID NO: 103 is provided which includes a) transforming a host cell having a naturally occurring amount of *CYP52A8B* protein with an increased copy number of a *CYP52A8B* gene that encodes the *CYP52A8B* protein having the amino acid sequence as set forth in SEQ ID NO: 103; and b) culturing the cell and thereby increasing expression of the protein compared with that of a host cell containing a naturally occurring copy number of the *CYP52A8B* gene.

A method for increasing production of a dicarboxylic acid is provided which includes a) providing a host cell having a naturally occurring number of *CYP52D4A* genes; b) increasing, in the host cell, the number of *CYP52D4A* genes which encode a *CYP52D4A* protein having the amino acid sequence as set forth in SEQ ID NO: 104; c) culturing the host cell in media containing an organic substrate which upregulates the *CYP52D4A* gene, to effect increased production of dicarboxylic acid.

A method for increasing the production of a *CYP52D4A* protein having an amino acid sequence as set forth in SEQ ID NO: 104 is provided which includes a) transforming a host cell having a naturally occurring amount of *CYP52D4A* protein with an increased copy number

of a *CYP52D4A* gene that encodes the *CYP52D4A* protein having the amino acid sequence as set forth in SEQ ID NO: 104; and b) culturing the cell and thereby increasing expression of the protein compared with that of a host cell containing a naturally occurring copy number of the *CYP52D4A* gene.

## BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1 is a schematic representation of cloning vector pTriplEx from Clontech™ Laboratories, Inc. Selected restriction sites within the multiple cloning site are shown.

Figure 2A is a map of the ZAP Express™ vector.

Figure 2B is a schematic representation of cloning phagemid vector pBK-CMV.

Figure 3 is a double stranded DNA sequence of a portion of the 5 prime coding region of the *CYP52A5A* gene (SEQ ID NO: 36).

Figure 4 is a diagrammatic representation of highly conserved regions of *CYP* and *CPR* gene protein sequences. Helix I represents the putative substrate binding site and HR2 represents the heme binding region. The FMN, FAD and NADPH binding regions are indicated below the *CPR* gene.

Figure 5 is a diagrammatic representation of the plasmid pHKM1 containing the truncated *CPRA* gene present in the pTriplEx vector. A detailed restriction map of only the sequenced region is shown at the top. The bar indicates the open reading frame. The direction of transcription is indicated by an arrow under the open reading frame.

Figure 6 is a diagrammatic representation of the plasmid pHKM4 containing the truncated *CPRA* gene present in the pTriplEx vector. A detailed restriction map of only the sequenced region is shown at the top. The bar indicates the open reading frame. The direction of transcription is indicated by an arrow under the open reading frame.

Figure 7 is a diagrammatic representation of the plasmid pHKM9 containing the *CPRB* gene (SEQ ID NO: 82) present in the pBK-CMV vector. A detailed restriction map of only the sequenced region is shown at the top. The bar indicates the open reading frame. The direction of transcription is indicated by an arrow under the open reading frame.

Figure 8 is a diagrammatic representation of the plasmid pHKM11 containing the *CYP52A1A* gene (SEQ ID NO: 85) present in the pBK-CMV vector. A detailed restriction map

of only the sequenced region is shown at the top. The bar indicates the open reading frame. The direction of transcription is indicated by an arrow under the open reading frame.

Figure 9 is a diagrammatic representation of the plasmid pHKM12 containing the *CYP52A8A* gene (SEQ ID NO: 92) present in the pBK-CMV vector. A detailed restriction map of only the sequenced region is shown at the top. The bar indicates the open reading frame. The direction of transcription is indicated by an arrow under the open reading frame.

Figure 10 is a diagrammatic representation of the plasmid pHKM13 containing the *CYP52D4A* gene (SEQ ID NO: 94) present in the pBK-CMV vector. A detailed restriction map of only the sequenced region is shown at the top. The bar indicates the open reading frame. The direction of transcription is indicated by an arrow under the open reading frame.

Figure 11 is a diagrammatic representation of the plasmid pHKM14 containing the *CYP52A2B* gene (SEQ ID NO: 87) present in the pBK-CMV vector. A detailed restriction map of only the sequenced region is shown at the top. The bar indicates the open reading frame. The direction of transcription is indicated by an arrow under the open reading frame.

Figure 12 is a diagrammatic representation of the plasmid pHKM15 containing the *CYP52A8B* gene (SEQ ID NO: 93) present in the pBK-CMV vector. A detailed restriction map of only the sequenced region is shown at the top. The bar indicates the open reading frame. The direction of transcription is indicated by an arrow under the open reading frame.

Figures 13A-13D show the complete DNA sequences including regulatory and coding regions for the *CPRA* gene (SEQ ID NO: 81) and *CPRB* gene (SEQ ID NO: 82) from *C. tropicalis* ATCC 20336. Figures 13A-13D show regulatory and coding region alignment of these sequences. Asterisks indicate conserved nucleotides. Bold indicates protein coding nucleotides; the start and stop codons are underlined.

Figure 14 shows the amino acid sequence of the *CPRA* (SEQ ID NO: 83) and *CPRB* (SEQ ID NO: 84) proteins from *C. tropicalis* ATCC 20336 and alignment of these amino acid sequences. Asterisks indicate residues which are not conserved.

Figures 15A-15M show the complete DNA sequences including regulatory and coding regions for the following genes from *C. tropicalis* ATCC 20366: *CYP52A1A* (SEQ ID NO: 85), *CYP52A2A* (SEQ ID NO: 86), *CYP52A2B* (SEQ ID NO: 87), *CYP52A3A* (SEQ ID NO: 88), *CYP52A3B* (SEQ ID NO: 89), *CYP52A5A* (SEQ ID NO. 90), *CYP52A5B* (SEQ ID NO: 91), *CYP52A8A* (SEQ ID NO: 92), *CYP52A8B* (SEQ ID NO: 93), and *CYP52D4*A (SEQ ID NO: 94).

Figures 15A-15M show regulatory and coding region alignment of these sequences. Asterisks indicate conserved nucleotides. Bold indicates protein coding nucleotides; the start and stop codons are underlined.

Figures 16A-16C show the amino acid sequences encoding the *CYP52A1A* (SEQ ID NO: 95), *CYP52A2A* (SEQ ID NO: 96), *CYP52A2B* (SEQ ID NO: 97), *CYP52A3A* (SEQ ID NO: 98), *CYP52A3B* (SEQ ID NO: 99), *CYP52A5A* (SEQ ID NO: 100), *CYP52A5B* (SEQ ID NO: 101), *CYP52A8A* (SEQ ID NO: 102), *CYP52A8B* (SEQ ID NO: 103) and *CYP52D4A* (SEQ ID NO. 104) proteins from *C. tropicalis* ATCC 20336. Asterisks indicate identical residues and dots indicate conserved residues.

Figure 17 is a diagrammatic representation of the pTAg PCR product cloning vector (commercially available from R&D Systems, Minneapolis, MN).

Figure 18 is a plot of the log ratio (U/C) of unknown target DNA product to competitor DNA product versus the concentration of competitor mRNA. The plot is used to calculate the target messenger RNA concentration in a quantitative competitive reverse transcription polymerase chain reaction (QC-RT-PCR).

Figure 19 is a graph showing the relative induction of *C. tropicalis* ATCC 20962 *CYP52A5A* (SEQ ID NO: 90) by the addition of the fatty acid substrate Emersol® 267 to the growth medium.

Figure 20 is a graph showing the induction of *C. tropicalis* ATCC 20962 *CYP52* and *CPR* genes by Emersol® 267. P450 genes *CYP52A3A* (SEQ ID NO: 88), *CYP52A3B* (SEQ ID NO: 89), and *CYP52D4A* (SEQ ID NO: 94) are expressed at levels below the detection level of the QC-RT-PCR assay.

Figure 21 is a scheme to integrate selected genes into the genome of *Candida tropicalis* strains and recovery of *URA3A* selectable marker.

Figure 22 is a schematic representation of the transformation of *C. tropicalis* H5343 ura3$^-$ with *CYP* and/or *CPR* genes. Only one *URA3* locus needs to be functional. There are a total of 6 possible *ura*3 targets (5*ura*3A loci-2 pox4 disruptions, 2 pox 5 disruptions, 1 *ura*3A locus; and 1 *ura*3B locus).

Figure 23 is the complete DNA sequence (SEQ ID NO: 105) encoding *URA3A* from *C. tropicalis* ATCC 20336 and the amino acid sequence of the encoded protein (SEQ ID NO: 106).

Figure 24 is a schematic representation of the plasmid pURAin, the base vector for integrating selected genes into the genome of *C. tropicalis*. The detailed construction of pURAin is described in the text.

Figure 25 is a schematic representation of the plasmid pNEB193 cloning vector (commercially available from New England Biolabs, Beverly, MA).

Figure 26 is a diagrammatic representation of the plasmid pPA15 containing the truncated *CYP52A2A* gene present in the pTriplEx vector. A detailed restriction map of only the sequenced region is shown at the top. The bar indicates the open reading frame. The direction of transcription is indicated by an arrow under the open reading frame.

Figure 27 is a schematic representation of pURA2in, the base vector is constructed in pNEB193 which contains the 8 bp recognition sequences for *Asc I*, *Pac I* and *Pme I*. *URA3A* (SEQ ID NO: 105) and *CYP52A2A* (SEQ ID NO: 86) do not contain these 8 bp recognition sites. *URA3A* is inverted so that the transforming fragment will attempt to recircularize prior to integration. An *Asc I/Pme I* fragment was used to transform H5343 $ura^-$.

Figure 28 shows a scheme to detect integration of *CYP52A2A* gene (SEQ ID NO: 86) into the genome of H5343 $ura^-$. In all cases, hybridization band intensity could reflect the number of integrations.

Figure 29 is a diagrammatic representation of the plasmid pPA57 containing the truncated *CYP52A3A* gene present in the pTriplEx vector. A detailed restriction map of only the sequenced region is shown at the top. The bar indicates the open reading frame. The direction of transcription is indicated by an arrow under the open reading frame.

Figure 30 is a diagrammatic representation of the plasmid pPA62 containing the truncated *CYP52A3B* gene present in the pTriplEx vector. A detailed restriction map of only the sequenced region is shown at the top. The bar indicates the open reading frame. The direction of transcription is indicated by an arrow under the open reading frame.

Figure 31 is a diagrammatic representation of the plasmid pPAL3 containing the truncated *CYP52A5A* gene present in the pTriplEx vector. A detailed restriction map of only the sequenced region is shown at the top. The bar indicates the open reading frame. The direction of transcription is indicated by an arrow under the open reading frame.

Figure 32 is a diagrammatic representation of the plasmid pPA5 containing the truncated *CYP52A5A* gene present in the pTriplEx vector. A detailed restriction map of only the

sequenced region is shown at the top. The bar indicates the open reading frame. The direction of transcription is indicated by an arrow under the open reading frame.

Figure 33 is a diagrammatic representation of the plasmid pPA18 containing the truncated *CYP52D4A* gene present in the pTriplEx vector. A detailed restriction map of only the sequenced region is shown at the top. The bar indicates the open reading frame. The direction of transcription is indicated by an arrow under the open reading frame.

Figure 34 is a graph showing the expression of *CYP52A1* (SEQ ID NO: 85), *CYP52A2* (SEQ ID NO: 86) and *CYP52A5* genes (SEQ ID NOS: 90 and 91) from *C. tropicalis* 20962 in a fermentor run upon the addition of amounts of the substrate oleic acid or tridecane in a spiking experiment.

Figure 35 depicts a scheme used for the extraction and analysis of diacids and monoacids from fermentation broths.

Figure 36 is a graph showing the induction of expression of *CYP52A1A*, *CYP52A2A* and *CYP52A5A* in a fermentor run upon addition of the substrate octadecane. No induction of *CYP52A3A* or *CYP52A3B* was observed under these conditions.

## DESCRIPTION OF THE PREFERRED EMBODIMENTS

Diacid productivity is improved according to the present invention by selectively increasing enzymes which are known to be important to the oxidation of organic substrates such as fatty acids composing the desired feed. According to the present invention, ten *CYP* genes and two *CPR* genes of *C. tropicalis* have been identified and characterized that relate to participation in the ω-hydroxylase complex catalyzing the first step in the ω-oxidation pathway. In addition, a novel quantitative competitive reverse transcription polymerase chain reaction (QC-RT-PCR) assay is used to measure gene expression in the fermentor under conditions of induction by one or more organic substrates as defined herein. Based upon QC-RT-PCR results, three *CYP* genes, *CYP52A1*, *CYP52A2* and *CYP52A5*, have been identified as being of greater importance for the ω-oxidation of long chain fatty acids. Amplification of the *CPR* gene copy number improves productivity. The QC-RT-PCR assay indicates that both *CYP* and *CPR* genes appear to be under tight regulatory control.

In accordance with the present invention, a method for discriminating members of a gene family by quantifying the amount of target mRNA in a sample is provided which

includes a) providing an organism containing a target gene; b) culturing the organism with an organic substrate which causes upregulation in the activity of the target gene; c) obtaining a sample of total RNA from the organism at a first point in time; d) combining at least a portion of the sample of the total RNA with a known amount of competitor RNA to form an RNA mixture, wherein the competitor RNA is substantially similar to the target mRNA but has a lesser number of nucleotides compared to the target mRNA; e) adding reverse transcriptase to the RNA mixture in a quantity sufficient to form corresponding target DNA and competitor DNA; (f) conducting a polymerase chain reaction in the presence of at least one primer specific for at least one substantially non-homologous region of the target DNA within the gene family, the primer also specific for the competitor DNA; g) repeating steps (c-f) using increasing amounts of the competitor RNA while maintaining a substantially constant amount of target RNA; h) determining the point at which the amount of target DNA is substantially equal to the amount of competitor DNA; i) quantifying the results by comparing the ratio of the concentration of unknown target to the known concentration of competitor; and j) obtaining a sample of total RNA from the organism at another point in time and repeating steps (d-i).

In addition, modification of existing promoters and/or the isolation of alternative promoters provides increased expression of *CYP* and *CPR* genes. Strong promoters are obtained from at least four sources: random or specific modifications of the *CYP52A2* promoter, *CYP52A5* promoter, *CYP52A1* promoter, the selection of a strong promoter from available *Candida* β-oxidation genes such as *POX4* and *POX5*, or screening to select another suitable *Candida* promoter.

Promoter strength can be directly measured using QT-RT-PCR to measure *CYP* and *CPR* gene expression in *Candida* cells isolated from fermentors. Enzymatic assays and antibodies specific for *CYP* and *CPR* proteins are used to verify that increased promoter strength is reflected by increased synthesis of the corresponding enzymes. Once a suitable promoter is identified, it is fused to the selected *CYP* and *CPR* genes and introduced into *Candida* for construction of a new improved production strain. It is contemplated that the coding region of the *CYP* and *CPR* genes can be fused to suitable promoters or other regulatory sequences which are well known to those skilled in the art.

In accordance with the present invention, studies on *C. tropicalis* ATCC 20336 have identified six unique *CYP* genes and four potential alleles. QC-RT-PCR analyses of cells

isolated during the course of the fermentation bioconversions indicate that at least three of the *CYP* genes are induced by fatty acids and at least two of the *CYP* genes are induced by alkanes. See Figure 34. Two of the *CYP* genes are highly induced indicating participation in the ω-hydroxylase complex which catalyzes the rate limiting step in the oxidation of fatty acids to the corresponding diacids.

The biochemical characterizations of each P450 enzyme herein is used to tailor the *C. tropicalis* host for optimal diacid productivity and is used to select P450 enzymes to be amplified based upon the fatty acid content of the feedstream. *CYP* gene(s) encoding P450 enzymes that have a low specific activity for the fatty acid or alkane substrate of choice are targeted for inactivation, thereby reducing the physiological load on the cell.

Since it has been demonstrated that *CPR* can be limiting in yeast systems, the removal of non-essential P450s from the system can free electrons that are being used by non-essential P450s and make them available to the P450s important for diacid productivity. Moreover, the removal of non-essential P450s can make available other necessary but potentially limiting components of the P450 system (i.e., available membrane space, heme and/or NADPH).

Diacid productivity is thus improved by selective integration, amplification, and over expression of *CYP* and *CPR* genes in the C. *tropicalis* production host.

It should be understood that host cells into which one or more copies of desired *CYP* and/or *CPR* genes have been introduced can be made to include such genes by any technique known to those skilled in the art. For example, suitable host cells include procaryotes such as *Bacillus sp.*, *Pseudomous sp.*, *Actinomycetes sp.*, *Eschericia sp.*, *Mycobacterium sp.*, and eukaryotes such as yeast, algae, insect cells, plant cells and and filamentous fungi. Suitable host cells are preferably yeast cells such as *Yarrowia*, *Bebaromyces*, *Saccharomyces*, *Schizosaccharomyces*, and *Pichia* and more preferably those of the *Candida* genus. Preferred species of *Candida* are *tropicalis*, *maltosa*, *apicola*, *paratropicalis*, *albicans*, *cloacae*, *guillermondii*, *intermedia*, *lipolytica*, *parapsilosis* and *zeylenoides*. Certain preferred stains of *Candida tropicalis* are listed in U.S. Patent No. 5,254,466, incorporated herein by refe*sform*

Vectors such as plasmids, phagemids, phages or cosmids can be us *cell a* or transfect suitable host cells. Host cells may also be transformed by introdu linear DNA vector(s) containing the desired gene sequence. Such linear D A *into the* advantageous when it is desirable to avoid introduction of non-native

cell. For example, DNA consisting of a desired target gene(s) flanked by DNA sequences which are native to the cell can be introduced into the cell by electroporation, lithium acetate transformation, spheroplasting and the like. Flanking DNA sequences can include selectable markers and/or other tools for genetic engineering.

A suitable organic substrate herein can be any organic compound that is biooxidizable to a mono- or polycarboxylic acid. Such a compound can be any saturated or unsaturated aliphatic compound or any carbocyclic or heterocyclic aromatic compound having at least one terminal methyl group, a terminal carboxyl group and/or a terminal functional group which is oxidizable to a carboxyl group by biooxidation. A terminal functional group which is a derivative of a carboxyl group may be present in the substrate molecule and may be converted to a carboxyl group by a reaction other than biooxidation. For example, if the terminal group is an ester that neither the wild-type *C. tropicalis* nor the genetic modifications described herein will allow hydrolysis of the ester functionality to a carboxyl group, then a lipase can be added during the fermentation step to liberate free fatty acids. Suitable organic substrates include, but are not limited to, saturated fatty acids, unsaturated fatty acids, alkanes, alkenes, alkynes and combinations thereof.

Alkanes are a type of saturated organic substrate which are useful herein. The alkanes can be linear or cyclic, branched or straight chain, substituted or unsubstituted. Particularly preferred alkanes are those having from about 4 to about 25 carbon atoms, examples of which include but are not limited to butane, hexane, octane, nonane, dodecane, tridecane, tetradecane, octadecane and the like.

Examples of unsaturated organic substrates which can be used herein include but are not limited to internal olefins such as 2-pentene, 2-hexene, 3-hexene, 9-octadecene and the like; unsaturated carboxylic acids such as 2-hexenoic acid and esters thereof, oleic acid and esters thereof including triglyceryl esters having a relatively high oleic acid content, erucic acid and esters thereof including triglyceryl esters having a relatively high erucic acid content, ricinoleic acid and esters thereof including triglyceryl esters having a relatively high ricinoleic acid content, leic acid and esters thereof including triglyceryl esters having a relatively high linoleic acid [illegible] unsaturated alcohols such as 3-hexen-1-ol, 9-octadecen-1-ol and the like; unsaturated o[illegible] such as 3-hexen-1-al, 9-octadecen-1-al and the like. In addition to the above, an [illegible]ate which can be used herein include alicyclic compounds having at least one

internal carbon-carbon double bond and at least one terminal methyl group, a terminal carboxyl group and/or a terminal functional group which is oxidizable to a carboxyl group by biooxidation. Examples of such compounds include but are not limited to 3,6-dimethyl, 1,4-cyclohexadiene; 3-methylcyclohexene; 3-methyl-1, 4-cyclohexadiene and the like.

Examples of the aromatic compounds that can be used herein include but are not limited to arenes such as o-, m-, p-xylene; o-, m-, p-methyl benzoic acid; dimethyl pyridine, and the like. The organic substrate can also contain other functional groups that are biooxidizable to carboxyl groups such as an aldehyde or alcohol group. The organic substrate can also contain other functional groups that are not biooxidizable to carboxyl groups and do not interfere with the biooxidation such as halogens, ethers, and the like.

Examples of saturated fatty acids which may be applied to cells incorporating the present *CYP* and *CPR* genes include caproic, enanthic, caprylic, pelargonic, capric, undecylic, lauric, myristic, pentadecanoic, palmitic, margaric, stearic, arachidic, behenic acids and combinations thereof. Examples of unsaturated fatty acids which may be applied to cells incorporating the present *CYP* and *CPR* genes include palmitoleic, oleic, erucic, linoleic, linolenic acids and combinations thereof. Alkanes and fractions of alkanes may be applied which include chain links from C12 to C24 in any combination. An example of a preferred fatty acid mixtures are Emersol® 267 and Tallow, both commercially available from Henkel Chemicals Group, Cincinnati, OH. The typical fatty acid composition of Emersol® 267 and Tallow is as follows:

| | TALLOW | E267 |
|---|---|---|
| C14:0 | 3.5% | 2.4% |
| C14:1 | 1.0% | 0.7% |
| C15:0 | 0.5% | -------- |
| C16:0 | 25.5% | 4.6% |
| C16:1 | 4.0% | 5.7% |
| C17:0 | 2.5% | -------- |
| C17:1 | -------- | 5.7% |
| C18:0 | 19.5% | 1.0% |
| C18:1 | 41.0% | 69.9% |
| C18:2 | 2.5% | 8.8% |

| | | |
|---|---|---|
| C18:3 | ——— | 0.3% |
| C20:0 | 0.5% | ——— |
| C20:1 | ——— | 0.9% |

The following examples are meant to illustrate but not to limit the invention. All relevant microbial strains and plasmids are described in Table 1 and Table 2, respectively.

**Table 1.** List of *Escherichia coli* and *Candida tropicalis* strains

| *E. Coli* STRAIN | GENOTYPE | SOURCE |
|---|---|---|
| XL1Blue-MRF' | *endA1, gyrA96, hsdR17, lac⁻, recA1, relA1, supE44, thi-1*, [F' *lacIqZ M15, proAB*, Tn*10*] | Stratagene, La Jolla, CA |
| BM25.8 | *SupE44, thi (lac-proAB)* [F' *traD36, proAB⁺, lacIqZ M15*] *λimm434 (kanR)P1 (camR) hsdR (rk12⁻ mk12⁻)* | Clontech, Palo Alto, CA |
| XLOLR | *(mcrA)183 (mcrCB-hsdSMR-mrr)173 endA1 thi-1 recA1 gyrA96 relA1 lac* [F'*proAB lacIqZ M15* Tn*10* (Tetr) Su⁻ (nonsuppressing λr(lambda resistant) | Stratagene, La Jolla, CA |

| *C. tropicalis* STRAIN | GENOTYPE | SOURCE |
|---|---|---|
| ATCC20336 | Wild-type | American Type Culture Collection, Rockville, MD |
| ATCC750 | Wild-type | American Type Culture Collection, Rockville, MD |
| ATCC 20962 | *ura3A/ura3B, pox4A::ura3A/pox4B::ura3A, pox5::ura3A/pox5::URA3A* | Henkel |
| H5343 ura- | *ura3A/ura3B, pox4A::ura3A/pox4B::ura3A, pox5::ura3A/pox5::URA3A, ura3-* | Henkel |
| HDC1 | *ura3A/ura3B, pox4A::ura3A/pox4B::ura3A, pox5::ura3A/pox5::URA3A, ura3::URA3A-CYP52A2A* | Henkel |
| HDC5 | *ura3A/ura3B, pox4A::ura3A/pox4B::ura3A, pox5::ura3A/pox5::URA3A, ura3::URA3A-CYP52A3A* | Henkel |
| HDC10 | *ura3A/ura3B, pox4A::ura3A/pox4B::ura3A, pox5::ura3A/pox5::URA3A, ura3::URA3A-CPRB* | Henkel |

| | | |
|---|---|---|
| HDC15 | *ura3A/ura3B, pox4A::ura3A/pox4B::ura3A, pox5::ura3A/pox5::URA3A, ura3::URA3A-CYP52A5A* | Henkel |
| HDC20 | *ura3A/ura3B, pox4A::ura3A/pox4B::ura3A, pox5::ura3A/pox5::URA3A, ura3::URA3A-CYP52A2A + CPR B* (*CYP* and *CPR* have opposite 5' to 3' orientation with respect to each other) | Henkel |
| HDC23 | *ura3A/ura3B, pox4A::ura3A/pox4B::ura3A, pox5::ura3A/pox5::URA3A, ura3::URA3A-CYP52A2A + CPR B* (*CYP* and *CPR* have same 5' to 3' orientation with respect to each other) | Henkel |

**Table 2.** List of plasmids isolated from genomic libraries and constructed for use in gene integrations.

| Plasmid | Base vector | Insert | Insert Size | Plasmid size | Description |
|---|---|---|---|---|---|
| pURAin | pNEB193 | *URA3A* | 1706 bp | 4399 bp | pNEB193 with the *URA3A* gene inserted in the *Asc*I - *Pme*I site, generating a *Pac*I site |
| pURA 2in | pURAin | *CYP52A2A* | 2230 bp | 6629 bp | pURAin containing a PCR *CYP52A2A* allele containing *Pac*I restriction sites |
| pURA REDB in | pURAin | *CPRB* | 3266 bp | 7665 bp | pURAin containing a PCR *CPRB* allele containing *Pac*I restriction sites |
| pHKM1 | pTriplEx | Truncated *CPRA* gene | Approx. 3.8 kb | Approx. 7.4 kb | A truncated *CPRA* gene obtained by first screening library containing the 5' untranslated region and 1.2 kb open reading frame |
| pHKM4 | PTriplEx | Truncated *CPRA* gene | Approx. 5 kb | Approx. 8.6 kb | A truncated *CPRA* gene obtained by screening second library containing the 3' untranslated region end sequence |
| pHKM9 | pBC-CMV | *CPRB* gene | Approx. 5.3 kb | Approx. 9.8 kb | *CPRB* allele isolated from the third library |
| pHKM11 | pBC-CMV | *CYP52A1A* | Approx. 5 kb | Approx. 9.5 kb | *CYP52A1A* isolated from the third library |
| pHKM12 | pBC-CMV | *CYP52A8A* | Approx. 7.5 kb | Approx. 12 kb | *CYP52A8A* isolated from the third library |
| pHKM13 | pBC-CMV | *CYP52D4A* | Approx. 7.3 kb | Approx. 11.8 kb | *CYP52D4A* isolated from the third library |

| | | | | | |
|---|---|---|---|---|---|
| pHKM14 | pBC-CMV | *CYP52A2B* | Approx. 6 kb | Approx. 10.5 kb | *CYP52A2B* isolated from the third library |
| pHKM15 | pBC-CMV | *CYP52A8B* | Approx. 6.6 kb | Approx. 11.1 kb | *CYP52A8B* isolated from the third library |
| pPAL3 | pTriplEx | *CYP52A5A* | 4.4 kb | Approx. 8.1 kb | *CYP52A5A* isolated from the 1st library |
| pPA5 | pTriplEx | *CYP52A5B* | 4.1 kb | Approx. 7.8 kb | *CYP52A5B* isolated from the 2nd library |
| pPA15 | pTriplEx | *CYP52A2A* | 6.0 kb | Approx. 9.7 kb | *CYP52A2A* isolated from the 2nd library |
| pPA57 | pTriplEx | *CYP52A3A* | 5.5 kb | Approx. 9.2 kb | *CYP52A3A* isolated from the 2nd library |
| pPA62 | pTriplEx | *CYP52A3B* | 6.0 kb | Approx. 9.7 kb | *CYP52A3B* isolated from the 2nd library |

## EXAMPLE 1

### Purification of Genomic DNA from *Candida tropicalis* ATCC 20336

### A. Construction of Genomic Libraries

50 ml of YEPD broth (see Chart) was inoculated with a single colony of *C. tropicalis* 20336 from YEPD agar plate and grown overnight at 30°C. 5 ml of the overnight culture was inoculated into 100 ml of fresh YEPD broth and incubated at 30°C for 4 to 5 hr with shaking. Cells were harvested by centrifugation, washed twice with sterile distilled water and resuspended in 4 ml of spheroplasting buffer (1 M Sorbitol, 50 mM EDTA, 14 mM mercaptoethanol) and incubated for 30 min at 37°C with gentle shaking. 0.5 ml of 2 mg/ml zymolyase (ICN Pharmaceuticals, Inc., Irvine, CA) was added and incubated at 37°C with gentle shaking for 30 to 60 min. Spheroplast formation was monitored by SDS lysis. Spheroplasts were harvested by brief centrifugation (4,000 rpm, 3 min) and were washed once with the spheroplast buffer without mercaptoethanol. Harvested spheroplasts were then suspended in 4 ml of lysis buffer (0.2 M Tris/pH 8.0, 50 mM EDTA, 1% SDS) containing 100 μg/ml RNase (Qiagen Inc., Chatsworth, CA) and incubated at 37°C for 30 to 60 min.

Proteins were denatured and extracted twice with an equal volume of chloroform/isoamyl alcohol (24:1) by gently mixing the two phases by hand inversions. The two phases were separated by centrifugation at 10,000 rpm for 10 min and the aqueous phase containing the high-molecular weight DNA was recovered. To the aqueous layer NaCl was added to a final concentration of 0.2 M and the DNA was precipitated by adding 2 vol of ethanol. Precipitated DNA was spooled with a clean glass rod and resuspended in TE buffer (10 mM

Tris/pH 8.0, 1 mM EDTA) and allowed to dissolve overnight at 4°C. To the dissolved DNA, RNase free of any DNase activity (Qiagen Inc., Chatsworth, CA) was added to a final concentration of 50 μg/ml and incubated at 37°C for 30 min. Then protease (Qiagen Inc., Chatsworth, CA) was added to a final concentration of 100 μg/ml and incubated at 55 to 60°C for 30 min. The solution was extracted once with an equal volume of phenol/chloroform/isoamyl alcohol (25:24:1) and once with equal volume of chloroform/isoamyl alcohol (24:1). To the aqueous phase 0.1 vol of 3 M sodium acetate and 2 volumes of ice cold ethanol (200 proof) were added and the high molecular weight DNA was spooled with a glass rod and dissolved in 1 to 2 ml of TE buffer.

### *B.* Genomic DNA Preparation for PCR Amplification of *CYP* and *CPR* Genes

Five 5 ml of YPD medium was inoculated with a single colony and grown at 30°C overnight. The culture was centrifuged for 5 min at 1200 x g. The supernatant was removed by aspiration and 0.5 ml of a sorbitol solution (0.9 M sorbitol, 0.1 M Tris-Cl pH 8.0, 0.1 M EDTA) was added to the pellet. The pellet was resuspended by vortexing and 1 μl of 2-mercaptoethanol and 50 μl of a 10 μg/ml zymolyase solution were added to the mixture. The tube was incubated at 37°C for 1 hr on a rotary shaker (200 rpm). The tube was then centrifuged for 5 min at 1200 x g and the supernatant was removed by aspiration. The protoplast pellet was resuspended in 0.5 ml 1x TE (10 mM Tris-Cl pH 8.0, 1 mM EDTA) and transferred to a 1.5 ml microcentrifuge tube. The protoplasts were lysed by the addition of 50 μl 10% SDS followed by incubation at 65°C for 20 min. Next, 200 μl of 5M potassium acetate was added and after mixing, the tube was incubated on ice for at least 30 min. Cellular debris was removed by centrifugation at 13,000 x g for 5 min. The supernatant was carefully removed and transferred to a new microfuge tube. The DNA was precipitated by the addition of 1 ml 100% (200 proof) ethanol followed by centrifugation for 5 min at 13,000 x g. The DNA pellet was washed with 1 ml 70 % ethanol followed by centrifugation for 5 min at 13,000 x g. After partially drying the DNA under a vacuum, it was resuspended in 200 μl of 1x TE. The DNA concentration was determined by ratio of the absorbance at 260 nm / 280 nm ($A_{260/280}$).

## EXAMPLE 2

### Construction of *Candida tropicalis* 20336 Gen mic Libraries

Three genomic libraries of *C. tropicalis* were constructed, two at Clontech Laboratories, Inc., (Palo Alto, CA) and one at Henkel Corporation (Cincinnati, OH).

### A. Clontech Libraries

The first Clontech library was made as follows: Genomic DNA was prepared from *C. tropicalis* 20336 as described above, partially digested with *Eco*RI and size fractionated by gel electrophoresis to eliminate fragments smaller than 0.6 kb. Following size fractionation, several ligations of the *Eco*RI genomic DNA fragments and lambda (λ) TriplEx™ vector (Figure 1) arms with *Eco*RI sticky ends were packaged into λ phage heads under conditions designed to obtain one million independent clones. The second genomic library was constructed as follows: Genomic DNA was digested partially with *Sau3A1* and size fractionated by gel electrophoresis. The DNA fragments were blunt ended using standard protocols as described, e.g., in Sambrook et al, *Molecular Cloning: A Laboratory Manual*, 2ed. Cold Spring Harbor Press, USA (1989), incorporated herein by reference. The strategy was to fill in the *Sau3A1* overhangs with Klenow polymerase (Life Technologies, Grand Island, NY) followed by digestion with S1 nuclease (Life Technologies, Grand Island, NY). After S1 nuclease digestion the fragments were end filled one more time with Klenow polymerase to obtain the final blunt-ended DNA fragments. *Eco*RI linkers were ligated to these blunt-ended DNA fragments followed by ligation into the λTriplEx vector. The resultant library contained approximately $2 \times 10^6$ independent clones with an average insert size of 4.5 kb.

### B. Henkel Library

The third genomic library was constructed at Henkel Corporation using λZAP Express™ vector (Stratagene, La Jolla, CA) (Figure 2). Genomic DNA was partially digested with *Sau3A1* and fragments in the range of 6 to 12 kb were purified from an agarose gel after electrophoresis of the digested DNA. These DNA fragments were then ligated to *Bam*HI digested λZAP Express™ vector arms according to manufacturers protocols. Three ligations were set up to obtain approximately $9.8 \times 10^5$ independent clones. All three libraries were pooled and amplified according to manufacturer instructions to obtain high-titre ($>10^9$ plaque

forming units/ml) stock for long-term storage. The titre of packaged phage library was ascertained after infection of *E. coli* XL1Blue-MRF' cells. *E. coli* XL1Blue-MRF' were grown overnight in either in LB medium or NZCYM (Chart) containing 10 mM $MgSO_4$ and 0.2% maltose at 37°C or 30°C, respectively with shaking. Cells were then centrifuged and resuspended in 0.5 to 1 volume of 10 mM $MgSO_4$. 200 µl of this *E. coli* culture was mixed with several dilutions of packaged phage library and incubated at 37°C for 15 min. To this mixture 2.5 ml of LB top agarose or NZCYM top agarose (maintained at 60°C ) (see Chart) was added and plated on LB agar or NCZYM agar (see Chart) present in 82 mm petri dishes. Phage were allowed to propagate overnight at 37°C to obtain discrete plaques and the phage titre was determined.

## EXAMPLE 3

### Screening of Genomic Libraries

Both λTriplEx™ and λZAP Express™ vectors are phagemid vectors that can be propagated either as phage or plasmid DNA (after conversion of phage to plasmid). Therefore, the genomic libraries constructed in these vectors can be screened either by plaque hybridization (screening of lambda form of library) or by colony hybridization (screening plasmid form of library after phage to plasmid conversion). Both vectors are capable of expressing the cloned genes and the main difference is the mechanism of excision of plasmid from the phage DNA. The cloning site in λTriplEx™ is located within a plasmid which is present in the phage and is flanked by *loxP* site (Figure 1). When λTriplEx™ is introduced into *E. coli* strain BM25.8 (supplied by Clontech), the *Cre* recombinase present in BM25.8 promotes the excision and circularization of plasmid pTriplEx from the phage λTriplEx™ at the *loxP* sites. The mechanism of excision of plasmid pBK-CMV from phage λZAP Express™ is different. It requires the assistance of a helper phage such as ExAssist™ (Stratagene) and an *E. coli* strain such as XLOR (Stratagene). Both pTriplEx and pBK-CMVcan replicate autonomously in *E. coli*.

## A. Screening Genomic Libraries (Plasmid Form)

### 1) Colony Lifts

A single colony of *E. coli* BM25.8 was inoculated into 5 ml of LB containing 50 µg/ml kanamycin, 10 mM $MgSO_4$ and 0.1% maltose and grown overnight at 31 °C, 250 rpm. To 200 µl of this overnight culture (~ 4 X $10^8$ cells) 1 µl of phage library (2 - 5 X $10^6$ plaque forming units) and 150 µl LB broth were added and incubated at 31 °C for 30 min after which 400 µl of LB broth was added and incubated at 31 °C , 225 rpm for 1 h. This bacterial culture was diluted and plated on LB agar containing 50 µg/ml ampicillin (Sigma Chemical Company, St. Louis, MO) and kanamycin (Sigma Chemical Company) to obtain 500 to 600 colonies/plate. The plates were incubated at 37°C for 6 to 7 hrs until the colonies became visible. The plates were then stored at 4°C for 1.5 h before placing a Colony/Plaque Screen™ Hybridization Transfer Membrane disc (DuPont NEN Research Products, Boston, MA) on the plate in contact with bacterial colonies. The transfer of colonies to the membrane was allowed to proceed for 3 to 5 min. The membrane was then lifted and placed on a fresh LB agar (see Chart) plate containing 200 µg/ml of chloramphenicol with the side exposed to the bacterial colonies facing up. The plates containing the membranes were then incubated at 37°C overnight in order to allow full development of the bacterial colonies. The LB agar plates from which colonies were initially lifted were incubated at 37°C overnight and stored at 4°C for future use. The following morning the membranes containing bacterial colonies were lifted and placed on two sheets of Whatman 3M (Whatman, Hillsboro, OR) paper saturated with 0.5 N NaOH and left at room temperature (RT) for 3 to 6 min to lyse the cells. Additional treatment of membranes was as described in the protocol provided by NEN Research Products.

### 2) DNA Hybridizations

Membranes were dried overnight before hybridizing to oligonucleotide probes prepared using a non-radioactive ECL™ 3'-oligolabelling and detection system from Amersham fe Sciences (Arlington Heights, IL). DNA labeling, prehybridization and hybridizations were med according to manufacturer's protocols. After hybridization, membranes were washed moom temperature in 5 X SSC, 0.1% SDS (in a volume equivalent to 2 ml/cm² of for 5 min each followed by two washes at 50°C in 1X SSC, 0.1% SDS (in a volume

equivalent to 2 ml/cm$^2$ of membrane) for 15 min each. The hybridization signal was then generated and detected with Hyperfilm ECL™ (Amersham) according to manufacturer's protocols. Membranes were aligned to plates containing bacterial colonies from which colony lifts were performed and colonies corresponding to positive signals on X-ray were then isolated and propagated in LB broth. Plasmid DNA's were isolated from these cultures and analyzed by restriction enzyme digestions and by DNA sequencing.

**B. Screening Genomic Libraries (Plaque Form)**

**1) λ Library Plating**

*E. coli* XL1Blue-MRF' cells were grown overnight in LB medium (25 ml) containing 10 mM $MgSO_4$ and 0.2% maltose at 37°C, 250 rpm. Cells were then centrifuged (2,200 x g for 10 min) and resuspended in 0.5 volumes of 10 mM $MgSO_4$. 500 µl of this *E. coli* culture was mixed with a phage suspension containing 25,000 amplified lambda phage particles and incubated at 37°C for 15 min. To this mixture 6.5 ml of NZCYM top agarose (maintained at 60°C) (see Chart) was added and plated on 80 - 100 ml NCZYM agar (see Chart) present in a 150 mm petridish. Phage were allowed to propagate overnight at 37°C to obtain discrete plaques. After overnight growth plates were stored in a refrigerator for 1-2 hr before plaque lifts were performed.

***2)* Plaque Lift and DNA Hybridizations**

Magna Lift™ nylon membranes (Micron Separations, Inc., Westborough, MA) were placed on the agar surface in complete contact with λ plaques and transfer of plaques to nylon membranes was allowed to proceed for 5 min at RT. After plaque transfer the membrane was placed on 2 sheets of Whatman 3M™ (Whatman, Hillsboro, OR) filter paper saturated with a 0.5 N NaOH, 1.0 M NaCl solution and left for 10 min at RT to denature DNA. Excess denaturing solution was removed by blotting briefly on dry Whatman 3M paper. Membranes were then transferred to 2 sheets of Whatman 3M™ paper saturated with 0.5 M Tris-HCl (pH 8.0), 1.5 M NaCl and left for 5 min to neutralize. Membranes were then briefly washed in 200 - 500 ml of 2 X SSC, dried by air and baked for 30 - 40 min at 80°C. The membranes were then probed with labelled DNA.

Membranes were prewashed with a 200 - 500 ml solution of 5 X SSC, 0.5% SDS, 1 mM EDTA (pH 8.0) for 1 - 2 hr at 42°C with shaking (60 rpm) to get rid of bacterial debris from the membranes. The membranes were prehybridized for 1 - 2 hr at 42°C with (in a volume equivalent to 0.125 - 0.25 ml/cm$^2$ of membrane) ECL Gold™ buffer (Amersham) containing 0.5 M NaCl and 5% blocking reagent. DNA fragments that were used as probes were purified from agarose gel using a QIAEX II™ gel extraction kit (Qiagen Inc., Chatsworth, CA) according to manufacturers protocol and labeled using an Amersham ECL™ direct nucleic acid labeling kit (Amersham). Labeled DNA (5 - 10 ng/ml hybridization solution) was added to the prehybridized membranes and the hybridization was allowed to proceed overnight. The following day membranes were washed with shaking (60 rpm) twice at 42°C for 20 min each time in (in a volume equivalent to 2 ml/cm$^2$ of membrane) a buffer containing either 0.1 (high stringency) or 0.5 (low stringency) X SSC, 0.4% SDS and 360 g/l urea. This was followed by two 5 min washes at room temperature in (in a volume equivalent to 2 ml/cm$^2$ of membrane) 2 X SSC. Hybridization signals were generated using the ECL™ nucleic acid detection reagent and detected using Hyperfilm ECL™ (Amersham).

Agar plugs which contained plaques corresponding to positive signals on the X-ray film were taken from the master plates using the broad-end of Pasteur pipet. Plaques were selected by aligning the plates with the x-ray film. At this stage, multiple plaques were generally taken. Phage particles were eluted from the agar plugs by soaking in 1 ml SM buffer (Sambrook et al., *supra*) overnight. The phage eluate was then diluted and plated with freshly grown *E. coli* XL1Blue-MRF' cells to obtain 100 - 500 plaques per 85 mm NCZYM agar plate. Plaques were transferred to Magna Lift nylon membranes as before and probed again using the same probe. Single well-isolated plaques corresponding to signals on X - ray film were picked by removing agar plugs and eluting the phage by soaking overnight in 0.5 ml SM buffer.

### C. Conversion of λ Clones to Plasmid Form

The lambda clones isolated were converted to plasmid form for further analysis. Conversion from the plaque to the plasmid form was accomplished by infecting the plaques into *E. coli* strain BM25.8. The *E. coli* strain was grown overnight at 31°C, 250 rpm in LB broth containing 10 mM $MgSO_4$ and 0.2% maltose until the $OD_{600}$ reached 1.1 - 1.4. Ten milliliters of the overnight culture was removed and mixed with 100 μl of 1 M $MgCl_2$. A 200 μl volume of

cells was removed, mixed with 150 µl of eluted phage suspension and incubated at 31 °C for 30 min. LB broth (400 µl) was added to the tube and incubation was continued at 31 °C for 1 hr with shaking, 250 rpm. 1 - 10 µl of the infected cell suspension was plated on LB agar containing 100 µg/ml ampicillin (Sigma, St. Louis, MO). Well-isolated colonies were picked and grown overnight in 5 ml LB broth containing 100 µg/ml ampicillin at 37°C, 250 rpm. Plasmid DNA was isolated from these cultures and analyzed. To convert the λZAP Express™ vector to plasmid form *E. coli* strains XL1Blue-MRF' and XLOR were used. The conversion was performed according to the manufacturer's (Stratagene) protocols for single-plaque excision.

## EXAMPLE 4

### Transformation of *C. tropicalis* H5343 ura⁻

#### A. Transformation of *C. tropicalis* H5343 by Electroporation

5 ml of YEPD was inoculated with *C. tropicalis* H5343 *ura*- from a frozen stock and incubated overnight on a New Brunswick shaker at 30°C and 170 rpm. The next day, 10 µl of the overnight culture was inoculated into 100 ml YEPD and growth was continued at 30°C, 170 rpm. The following day the cells were harvested at an $OD_{600}$ of 1.0 and the cell pellet was washed one time with sterile ice-cold water. The cells were resuspended in ice-cold sterile 35 % Polyethylene glycol (4,000 MW) to a density of $5x10^8$ cells/ml. A 0.1 ml volume of cells were utilized for each electroporation. The following electroporation protocol was followed: 1.0 $\mu$g of transforming DNA was added to 0.1 ml cells, along with 5 $\mu$g denatured, sheared calf thymus DNA and the mixture was allowed to incubate on ice for 15 min. The cell solution was then transferred to an ice-cold 0.2 cm electroporation cuvette, tapped to make sure the solution was on the bottom of the cuvette and electroporated. The cells were electroporated using an Invitrogen electroporator (Carlsbad, CA) at 450 Volts, 200 Ohms and 250 $\mu$F. Following electroporation, 0.9 ml SOS media (1M Sorbitol, 30% YEPD, 10 mM $CaCl_2$) was added to the suspension. The resulting culture was grown for 1 hr at 30°C, 170 rpm. Following the incubation, the cells were pelleted by centrifugation at 1500 x g for 5 min. The electroporated cells were resuspended in 0.2 ml of 1M sorbitol and plated on synthetic complete media minus uracil (SC - uracil) (Nelson, *supra*). In some cases the electroporated cells were

plated directly onto SC - uracil. Growth of transformants was monitored for 5 days. After three days, several transformants were picked and transferred to SC-uracil plates for genomic DNA preparation and screening.

**B. Transformation of *C. tropicalis* Using Lithium Acetate**

The following protocol was used to transform *C. tropicalis* in accordance with the procedures described in *Current Protocols in Molecular Biology*, Supplement 5, 13.7.1 (1989), incorporated herein by reference.

5 ml of YEPD was inoculated with *C. tropicalis* H5343 *ura-* from a frozen stock and incubated overnight on a New Brunswick shaker at 30°C and 170 rpm. The next day, 10 μl of the overnight culture was inoculated into 50 ml YEPD and growth was continued at 30°C, 170 rpm. The following day the cells were harvested at an $OD_{600}$ of 1.0. The culture was transferred to a 50 ml polypropylene tube and centrifuged at 1000 X g for 10 min. The cell pellet was resuspended in 10 ml sterile TE (10mM Tris-Cl and 1mM EDTA, pH 8.0). The cells were again centrifuged at 1000 X g for 10 min and the cell pellet was resuspended in 10 ml of a sterile lithium acetate solution [LiAc ( 0.1 M lithium acetate, 10 mM Tris-Cl, pH 8.0, 1 mM EDTA)]. Following centrifugation at 1000 X g for 10 min., the pellet was resuspended in 0.5 ml LiAc. This solution was incubated for one hour at 30°C while shaking gently at 50 rpm. A 0.1 ml aliquot of this suspension was incubated with 5 μg of transforming DNA at 30°C with no shaking for 30 min. A 0.7 ml PEG solution (40 % wt/vol polyethylene glycol 3340, 0.1 M lithium acetate, 10 mM Tris-Cl, pH 8.0, 1 mM EDTA) was added and incubated at 30°C for 45 min. The tubes were then placed at 42°C for 5 min. A 0.2 ml aliquot was plated on synthetic complete media minus uracil (SC - uracil) (Kaiser et al. *Methods in Yeast Genetics*, Cold Spring Harbor Laboratory Press, USA, 1994, incorporated herein by reference). Growth of transformants was monitored for 5 days. After three days, several transformants were picked and transferred to SC-uracil plates for genomic DNA preparation and screening.

## EXAMPLE 5

**Plasmid DNA Isolation**

Plasmid DNA were isolated from *E. coli* cultures using Qiagen plasmid isolation kit (Qiagen Inc., Chatsworth, CA) according to manufacturer's instructions.

## EXAMPLE 6

### DNA Sequencing and Analysis

DNA sequencing was performed at Sequetech Corporation (Mountain View, CA) using Applied Biosystems automated sequencer (Perkin Elmer, Foster City, CA). DNA sequences were analyzed with MacVector and GeneWorks software packages (Oxford Molecular Group, Campbell, CA).

## EXAMPLE 7

### PCR Protocols

PCR amplification was carried out in a Perkin Elmer Thermocycler using the Ampli*Taq*Gold enzyme (Perkin Elmer Cetus, Foster City, CA) kit according to manufacturer's specifications. Following successful amplification, in some cases, the products were digested with the appropriate enzymes and gel purified using QiaexII (Qiagen, Chatsworth, CA) as per manufacturer instructions. In specific cases the Ultma *Taq* polymerase (Perkin Elmer Cetus, Foster City, CA) or the Expand Hi-Fi *Taq* polymerase (Boehringer Mannheim, Indianapolis, IN) were used per manufacturer's recommendations or as defined in Table 3.

**Table 3.** PCR amplification conditions used with different primer combinations.

| PRIMER COMBINATION | *Taq* | TEMPLATE DENATURING CONDITION | ANNEALING TEMP/TIME | EXTENSION TEMP/TIME | CYCLE Number |
|---|---|---|---|---|---|
| 3674-41-1/ 41-2/ 41-4 + 3674-41-4 | Ampli-*Taq* Gold | 94 C/30 sec | 55 C/30 sec | 72 C/1 min | 30 |
| URA Primer 1a<br>URA Primer 1b | Ampli-*Taq* Gold | 95 C/1 min | 70 C/1 min | 72 C/2 min | 35 |
| URA Primer 2a<br>URA Primer 2b | Ampli-*Taq* Gold | 95 C/1 min | 70 C/1 min | 72 C/2 min | 35 |
| *CYP2A*#1<br>*CYP2A*#2 | Ampli-*Taq* Gold | 95 C/1 min | 70 C/1 min | 72 C/2 min | 35 |
| *CYP3A*#1<br>*CYP3A*#2 | Ultma *Taq* | 95 C/1 min | 70 C/1 min | 72 C/1 min | 30 |
| *CPR* B#1<br>*CPR* B#2 | Expand Hi-Fi *Taq* | 94 C/15 sec<br>94 C/15 sec | 50 C/30 sec<br>50 C/30 sec | 68 C/3 min<br>68 C/3 min<br>+20 sec/cycle | 10<br>15 |

| | | | | | |
|---|---|---|---|---|---|
| *CYP*5A#1<br>*CYP*5A#2 | Expand<br>Hi-Fi<br>*Taq* | 94 C/15 sec<br>94 C/15 sec | 50 C/30 sec<br>50 C/30 sec | 68 C/3 min<br>68 C/3 min<br>+20 sec/cycle | 10<br>15 |

Table 4 below contains a list of primers (SEQ ID NOS: 1-35) used for PCR amplification to construct gene integration vectors or to generate probes for gene detection and isolation.

**Table 4.** Primer table for PCR amplification to construct gene integration vectors, to generate probes for gene isolation and detection and to obtain DNA sequence of constructs. (A- deoxyadenosine triphosphate [dATP], G- deoxyguanosine triphosphate [dGTP], C- deoxycytosine triphosphate [dCTP], T- deoxythymidine triphosphate [dTTP], Y- dCTP or dTTP, R- dATP or dGTP, W- dATP or dTTP, M- dATP or dCTP, N- dATP or dCTP or dGTP or dTTP).

| Target gene(s) | Patent Primer Name | Lab Primer Name | Sequence (5' to 3') | PCR Product Size |
|---|---|---|---|---|
| *CYP52A2A* | CYP2A#1 | 3659-72M | *CCTTAATTAA*ATGCACGAAGCGGAGA TAAAAG<br>(SEQ ID NO: 1) | 2230 bp |
| | CYP2A#2 | 3659-72N | *CCTTAATTAA*GCATAAGCTTGCTCGAG TCT<br>(SEQ ID NO: 2) | |
| *CYP52A3A* | CYP3A#1 | 3659-72O | *CCTTAATTAA*ACGCAATGGGAACATG GAGTG<br>(SEQ ID NO: 3) | 2154 bp |
| | CYP3A#2 | 3659-72P | *CCTTAATTAA*TCGCACTACGGTTATTG GTATCAG<br>(SEQ ID NO: 4) | |
| *CYP52A5A* | CYP5A#1 | 3659-72K | *CCTTAATTAA*TCAAAGTACGTTCAGGC GG<br>(SEQ ID NO: 5) | 3298 bp |
| | CYP5A#2 | 3659-72L | *CCTTAATTAA*GGCAGACAACAACTTG GCAAAGTC<br>(SEQ ID NO: 6) | |
| *CPRB* | CPRB#1 | 3698-20A | *CCTTAATTAA*GAGGTCGTTGGTTGAGT TTTC<br>(SEQ ID NO: 7) | 3266 bp |
| | CPRB#2 | 3698-20B | *CCTTAATTAA*TTGATAATGACGTTGCG GG<br>(SEQ ID NO: 8) | |
| *URA3A* | URA Primer 1a | 3698-7C | *AGGCGCGCC*GGAGTCCAAAAAGACC AACCTCTG<br>(SEQ ID NO: 9) | 956 bp |
| | URA Primer 1b | 3698-7D | *CCTTAATTAA*TACGTGGATACCTTCAA GCAAGTG<br>(SEQ ID NO: 10) | |

| | | | | |
|---|---|---|---|---|
| *URA3A* | URA Primer 2a | 3698-7A | *CCTTAATTAA*GCTCACGAGTTTTGGGATTTTCGAG (SEQ ID NO: 11) | 750 bp |
| | URA Primer 2b | 3698-7B | *GGGTTTAAAC*CGCAGAGGTTGGTCTTTTTGGACTC (SEQ ID NO: 12) | |
| | | | | |
| | | | *GGGTTTAAAC* - *Pme* I restriction site (SEQ ID NO: 13) | |
| | | | *AGGCGCGCC* - *Asc*I restriction site (SEQ ID NO: 14) | |
| | | | *CCTTAATTAA* - *Pac*I restriction site (SEQ ID NO: 15) | |
| | | | | |
| *CPR* | FMN1 | 3674-41-1 | TCYCAAACWGGTACWGCWGAA (SEQ ID NO: 16) | |
| *CPR* | FMN2 | 3674-41-2 | GGTTTGGGTAAYTCWACTTAT (SEQ ID NO: 17) | |
| *CPR* | FAD | 3674-41-3 | CGTTATTAYTCYATTTCTTC (SEQ ID NO: 18) | |
| *CPR* | NADPH | 3674-41-4 | GCMACACCRGTACCTGGACC (SEQ ID NO: 19) | |
| *CPR* | PRK1.F3 | PRK1.F3 | ATCCCAATCGTAATCAGC (SEQ ID NO: 20) | |
| *CPR* | PRK1.F5 | PRK1.F5 | ACTTGTCTTCGTTTAGCA (SEQ ID NO: 21) | |
| *CPR* | PRK4.R20 | PRK4.R20 | CTACGTCTGTGGTGATGC (SEQ ID NO: 22) | |
| *CYP* | UCup1 | UCup1 | CGNGAYACNACNGCNGG (SEQ ID NO: 23) | |
| *CYP* | UCup2 | UCup2 | AGRGAYACNACNGCNGG (SEQ ID NO: 24) | |
| *CYP* | UCdown1 | UCdown1 | AGNGCRAAYTGYTGNCC (SEQ ID NO: 25) | |
| *CYP* | UCdown2 | UCdown2 | YAANGCRAAYTGYTGNCC (SEQ ID NO: 26) | |
| *CYP* | HemeB1 | HemeB1 | ATTCAACGGTGGTCCAAGAATCTGTTTGG (SEQ ID NO: 27) | |
| *CYP* | 2,3,5P | 2,3,5P | GAGCTATGTTGAGACCACAGTTTGC (SEQ ID NO: 28) | |
| *CYP* | 2,3,5M | 2,3,5M | CTTCAGTTAAAGCAAATTGTTTGGCC (SEQ ID NO: 29) | |
| pTriplEx vector | Triplex5' | Triplex5' | CTCGGGAAGCGCGCCATTGTGTTGG (SEQ ID NO: 30) | |
| pTriplEx vector | Triplex3' | Triplex3' | TAATACGACTCACTATAGGGCGAATTGGC (SEQ ID NO: 31) | |
| *CYP* | Cyp52a | Cyp52a | TGRYTCAAACCATCTYTCTGG (SEQ ID NO: 32) | |
| *CYP* | Cyp52b | Cyp52b | GGACCGGCGTTAAAGGG (SEQ ID NO: 33) | |
| *CYP* | Cyp52c | Cyp52c | CATAGTCGWATYATGCTTAGACC (SEQ ID NO: 34) | |
| *CYP* | Cyp52d | Cyp52d | GGACCACCATTGAATGG (SEQ ID NO: 35) | |

## EXAMPLE 8

### Yeast C lony PCR Procedure for Confirmati n of Gene Integration into the Genome f *C. tropicalis*

Single yeast colonies were removed from the surface of transformation plates, suspended in 50 $\mu$l of spheroplasting buffer (50mM KCl, 10mM Tris-HCl, pH 8.3, 1.0 mg/ml Zymolyase, 5% glycerol) and incubated at 37°C for 30 min. Following incubation, the solution was heated for 10 min at 95°C to lyse the cells. Five $\mu$l of this solution was used as a template in PCR. Expand Hi-Fi *Taq* polymerase (Boehringer Mannheim, Indianapolis, IN) was used in PCR coupled with a gene-specific primer (gene to be integrated) and a *URA3* primer. If integration did occur, amplification would yield a PCR product of predicted size confirming the presence of an integrated gene.

## EXAMPLE 9

### Fermentation Method for Gene Induction Studies

A fermentor was charged with a semi-synthetic growth medium having the composition 75 g/l glucose (anhydrous), 6.7 g/l Yeast Nitrogen Base (Difco Laboratories), 3 g/l yeast extract, 3 g/l ammonium sulfate, 2 g/l monopotassium phosphate, 0.5 g/l sodium chloride. Components were made as concentrated solutions for autoclaving then added to the fermentor upon cooling: final pH approximately 5.2. This charge was inoculated with 5-10% of an overnight culture of *C. tropicalis* ATCC 20962 prepared in YM medium (Difco Laboratories) as described in the methods of Examples 17 and 20 of US Patent 5,254,466, which is incorporated herein by reference. *C. tropicalis* ATCC 20962 is a POX 4 and POX 5 disrupted *C. tropicalis* ATCC 20336. Air and agitation were supplied to maintain the dissolved oxygen at greater than about 40% of saturation versus air. The pH was maintained at about 5.0 to 8.5 by the addition of 5N caustic soda on pH control. Both a fatty acid feedstream (commercial oleic acid in this example) having a typical composition: 2.4% $C_{14}$; 0.7% $C_{14:1}$; 4.6% $C_{16}$; 5.7% $C_{16:1}$; 5.7% $C_{17:1}$; 1.0% $C_{18}$; 69.9% $C_{18:1}$; 8.8% $C_{18:2}$; 0.30% $C_{18:3}$; 0.90% $C_{20:1}$ and a glucose co-substrate feed were added in a feedbatch mode beginning near the end of exponential growth. Caustic was added on pH control during the bioconversion of fatty acids to diacids to maintain the pH in the desired range. Typically, samples for gene induction studies were collected just prior to starting the fatty acid feed and over the first 10 hours of bioconversion. Determination of fatty acid and diacid

content was determined by a standard methyl ester protocol using gas liquid chromatography (GLC). Gene induction was measured using the QC-RT-PCR protocol described in this application.

## EXAMPLE 10

### RNA Preparation

The first step of this protocol involves the isolation of total cellular RNA from cultures of *C. tropicalis*. The cellular RNA was isolated using the Qiagen RNeasy Mini Kit (Qiagen Inc., Chatsworth, CA) as follows: 2 ml samples of *C. tropicalis* cultures were collected from the fermentor in a standard 2 ml screw capped Eppendorf style tubes at various times before and after the addition of the fatty acid or alkane substrate. Cell samples were immediately frozen in liquid nitrogen or a dry-ice/alcohol bath after their harvesting from the fermentor. To isolate total RNA from the samples, the tubes were allowed to thaw on ice and the cells pelleted by centrifugation in a microfuge for 5 minutes (min) at 4°C and the supernatant was discarded while keeping the pellet ice-cold. The microfuge tubes were filled 2/3 full with ice-cold Zirconia/Silica beads (0.5 mm diameter, Biospec Products, Bartlesville, OK) and the tube filled to the top with ice-cold RLT* lysis buffer (* buffer included with the Qiagen RNeasy Mini Kit). Cell rupture was achieved by placing the samples in a mini bead beater (Biospec Products, Bartlesville, OK) and immediately homogenized at full speed for 2.5 min. The samples were allowed to cool in a ice water bath for 1 minute and the homogenization/cool process repeated two more times for a total of 7.5 min homogenization time in the beadbeater. The homogenized cells samples were microfuged at full speed for 10 min and 700 µl of the RNA containing supernatant removed and transferred to a new eppendorf tube. 700 µl of 70% ethanol was added to each sample followed by mixing by inversion. This and all subsequent steps were performed at room temperature. Seven hundred microliters of each ethanol treated sample were transferred to a Qiagen RNeasy spin column, followed by centrifugation at 8,000 x g for 15 sec. The flow through was discarded and the column reloaded with the remaining sample (700 µl) and re-centrifuged at 8,000 x g for 15 sec. The column was washed once with 700 µl of buffer RW1*, and centrifuged at 8,000 x g for 15 sec and the flow through discarded. The column was placed in a new 2 ml collection tube and washed with 500 µl of RPE* buffer and the flow through discarded. The RPE* wash was repeated with centrifugation at 8,000 x g for 2 min and the flow through

discarded. The spin column was transferred to a new 1.5 ml collection tube and 100 μl of RNase free water added to the column followed by centrifugation at 8,000 x g for 15 seconds. An additional 75 μl of RNase free water was added to the column followed by centrifugation at 8,000 x g for 2 min. RNA eluted in the water flow through was collected for further purification.

The RNA eluate was then treated to remove contaminating DNA. Twenty microliters of 10X DNase I buffer (0.5 M tris (pH 7.5), 50 mM $CaCl_2$, 100 mM $MgCl_2$), 10 μl of RNase-free DNase I (2 Units/μl, Ambion Inc., Austin, Texas) and 40 units Rnasin (Promega Corporation, Madison, Wisconsin) were added to the RNA sample. The mixture was then incubated at 37°C for 15 to 30 min. Samples were placed on ice and 250 μl Lysis buffer RLT* and 250 μl ethanol (200 proof) added. The samples were then mixed by inversion. The samples were transferred to Qiagen RNeasy spin columns and centrifuged at 8,000 x g for 15 sec and the flow through discarded. Columns were placed in new 2 ml collection tubes and washed twice with 500 μl of RPE* wash buffer and the flow through discarded. Columns were transferred to new 1.5 ml eppendorf tubes and RNA was eluated by the addition of 100 μl of DEPC treated water followed by centrifugation at 8,000 x g for 15 sec. Residual RNA was collected by adding an additional 50 μl of RNase free water to the spin column followed by centrifugation at full speed for 2 min. 10 μl of the RNA preparation was removed and quantified by the ($A_{260/280}$) method. RNA was stored at

-70°C. Yields were found to be 30-100 μg total RNA per 2.0 ml of fermentation broth.

## EXAMPLE 11

### Quantitative Competitive Reverse Transcription Polymerase Chain Reaction (QC-RT-PCR) Protocol

QC-RT-PCR is a technique used to quantitate the amount of a specific RNA in a RNA sample. This technique employs the synthesis of a specific DNA molecule that is complementary to an RNA molecule in the original sample by reverse transcription and its subsequent amplification by polymerase chain reaction. By the addition of various amounts of a competitor RNA molecule to the sample one can determine the concentration of the RNA molecule of interest (in this case the mRNA transcripts of the *CYP* and *CPR* genes). The levels of specific mRNA transcripts were assayed over time in response to the addition of fatty acid

and/or alkane substrates to the growth medium of fermentation grown *C. tropicalis* cultures for the identification and characterization of the genes involved in the oxidation of these substrates. This approach can be used to identify the *CYP* and *CPR* genes involved in the oxidation of any given substrate based upon their transcriptional regulation.

**A. Primer Design**

The first requirement for QC-RT-PCR is the design of the primer pairs to be used in the reverse transcription and subsequent PCR reactions. These primers need to be unique and specific to the gene of interest. As there is a family of genetically similar *CYP* genes present in *C. tropicalis* 20336, care had to be taken to design primer pairs that would be discriminating and only amplify the gene of interest, in this example the *CYP52A5* gene. In this manner, unique primers directed to substantially non-homologous (aka variable) regions within target members of a gene family are constructed. What constitutes substantially non-homologous regions is determined on a case by case basis. Such unique primers should be specific enough to anneal the non-homologous region of the target gene without annealing to other non-target members of the gene family. By comparing the known sequences of the members of a gene family, non-homologous regions are identified and unique primers are constructed which will anneal to those regions. It is contemplated that non-homologous regions herein would typically exhibit less than about 85% homology but can be more homologous depending on the positions which are conserved and stringency of the reaction. After conducting PCR, it may be helpful to check the reaction product to assure it represents the unique target gene product. If not, the reaction conditions can be altered in terms of stringency to focus the reaction to the desired target. Alternatively a new primer or new non-homologous region can be chosen. Due to the high level of homology between the genes of the *CYP52A* family, the most variable 5 prime region of the *CYP52A5* coding sequence was targeted for the design of the primer pairs. In Figure 3, a portion of the 5 prime coding region for the *CYP52A5*A (SEQ ID NO: 36) allele of *C. tropicalis* 20336 is shown. The boxed sequences in Figure 3 are the sequences of the forward and backwards primers (SEQ ID NOS: 47 and 48) used to quantitate expression of both alleles of this gene. The actual reverse primer (SEQ ID NO: 48) contains one less adenine than that shown in Figure 3. Primers used to measure the expression of specific *C. tropicalis* 20336 genes using the QC-RT-PCR protocol are listed in Table 5 (SEQ ID NOS: 37-58).

**Table 5.** Primer used to measure *C. tropicalis* gene expression in the QC-RT-PCR reactions.

| Primer Name | Direction | Target | Sequence |
|---|---|---|---|
| 3737-89F | F | *CYP52A1A* | CCGATGAAGTTTTCGACGAGTACCC (SEQ ID NO: 37) |
| 3737-89B | B | *CYP52A1A* | AAGGCTTTAACGTGTCCAATCTGGTC (SEQ ID NO: 38) |
| alk2aF1 | F | *CYP52A2A* | ATTATCGCCACATACTTCACCAAATGG (SEQ ID NO: 39) |
| alk2aB5 | B | *CYP52A2A* | CGAGATCGTGGATACGCTGGAGTG (SEQ ID NO: 40) |
| 7581-178-3 | F | *CYP52A3A* | GCCACTCGGTAACTTTGTCAGGGAC (SEQ ID NO: 41) |
| 7581-178-4 | B | *CYP52A3A* | CATTGAACTGAGTAGCCAAAACAGCC (SEQ ID NO: 42) |
| 3737-50F | F | *CYP52A3A* & *CYP52A3B* | CCTACGTTTGGTATCGCTACTCCGTTG (SEQ ID NO: 43) |
| 3737-50B | B | *CYP52A3A* & *CYP52A3B* | TTTCCAGCCAGCACCGTCCAAG (SEQ ID NO: 44) |
| 3737-175F | F | *CYP52D4A* | GCAGAGCCGATCTATGTTGCGTCC (SEQ ID NO: 45) |
| 3737-175B | B | *CYP52D4A* | TCATTGAATGCTTCCAGGAACCTCG (SEQ ID NO: 46) |
| 7581-97-F | F | *CYP52A5A*& *CYP52A5B* | AAGAGGGCAGGGCTCAAGAG (SEQ ID NO: 47) |
| 7581-97-M | B | *CYP52A5A*& *CYP52A5B* | TCCATGTGAAGATCCCATCAC (SEQ ID NO: 48) |
| 4P-2 | F | *CYP52A8A* | CTTGAAGGCCGTGTTGAACG (SEQ ID NO: 49) |
| 4M-1 | B | *CYP52A8A* | CAGGATTTGTCTGAGTTGCCG (SEQ ID NO: 50) |
| 3737-52F | F | *POX4A* & *POX4B* | CCATTGCCTTGAGATACGCCATTGGTAG (SEQ ID NO: 51) |
| 3737-52B | B | *POX4A* & *POX4B* | AGCCTTGGTGTCGTTCTTTTCAACGG (SEQ ID NO: 52) |
| 3737-53F | F | *POX5A* | TTGGGTTTGTTTGTTTCCTGTGTCCG (SEQ ID NO: 53) |
| 3737-53B | B | *POX5A* | CCTTTGACCTTCAATCTGGCGTAGACG (SEQ ID NO: 54) |
| F33 | F | *CPRA* | GGTTTGCTGAATACGCTGAAGGTGATG (SEQ ID NO: 55) |
| B63 | B | *CPRA* | TGGAGCTGAACAACTCTCTCGTCTCGG (SEQ ID NO: 56) |
| 3737-133F | F | *CPRA* & *CPRB* | TTCCTCAACACGGACAGCGG (SEQ ID NO: 57) |
| 3737-133B | B | *CPRA* & *CPRB* | AGTCAACCAGGTGTGGAACTCGTC (SEQ ID NO: 58) |

F=Forward B=Backward

### B. Design and Synthesis of the Competitor DNA Template

The competitor RNA is synthesized *in vitro* from a competitor DNA template that has the T7 polymerase promoter and preferably carries a small deletion of e.g., about 10 to 25 nucleotides relative to the native target RNA sequence. The DNA template for the *in-vitro* synthesis of the competitor RNA is synthesized using PCR primers that are between 46 and 60 nucleotides in length. In this example, the primer pairs for the synthesis of the *CYP52A5* competitor DNA are shown in Tables 6 and 7 (SEQ ID NOS: 59 AND 60).

**Table 6.** Forward and Reverse primers used to synthesize the competitor RNA template for the QC-RT-PCR measurement of *CYP52A5A* gene expression.

| | | |
|---|---|---|
| Forward Primer | *CYP52A5A* | GGATCCTAATACGACTCACTATAGGGAGGA AGAGGGCAGGGCTCAAGAG (SEQ ID NO: 59) |
| Reverse Primer | *CYP52A5A* | TCCATGTGAAGATCCCATCACGAGTGTGCC TCTTGCCCAAAG (SEQ ID NO: 60) |

**Table 7.** Primers for the synthesis of the QC-RT-PCR competitor RNA templates

| Primer Name | Direction | Target | Sequence 5'-3' |
|---|---|---|---|
| 3737-89C | F | *CYP52A1A* | GGATCCTAATACGACTCACTATAGGGAGGCCGATG AAGTTTTCGACGAGTACCC (SEQ ID NO: 61) |
| 3737-89D | B | *CYP52A1A* | AAGGCTTTAACGTGTCCAATCTGGTC AACATAGCTCTGGAGTGCTTCCAACC (SEQ ID NO: 62) |
| 7581-137-A | F | *CYP52A2A* | GGATCCTAATACGACTCACTATAGGGAGGATTATC GCCACATACTTCACCAAATGG (SEQ ID NO: 63) |
| 7581-137-B | B | *CYP52A2A* | CGAGATCGTGGATACGCTGGAGTGCGTCGCTCTTC TTCTTCAACAATTCAAG (SEQ ID NO: 64) |
| 7581-137-D | B | *CYP52A3A* | CATTGAACTGAGTAGCCAAAACAGCCCATGGTTTC AATCAATGGGAGGC (SEQ ID NO: 65) |
| 7581-137-C | F | *CYP52A3A* | GGATCCTAATACGACTCACTATAGGGAGGGCCACT CGGTAACTTTGTCAGGGAC (SEQ ID NO: 66) |

| | | | |
|---|---|---|---|
| 3737-50-D | F | *CYP52A3A* & *CYP52A3B* | GGATCCTAATACGACTCACTATAGGGAGGCCTACG TTTGGTATCGCTACTCCGTTG (SEQ ID NO: 67) |
| 3737-50-C | B | *CYP52A3A* & *CYP52A3B* | TTTCCAGCCAGCACCGTCCAAGCAACAAGGAGTAC AAGAAATCGTGTC (SEQ ID NO: 68) |
| 3737-175C | F | *CYP52D4A* | GGATCCTAATACGACTCACTATAGGGAGGGCAGAG CCGATCTATGTTGCGTCC (SEQ ID NO: 69) |
| 3737-175D | B | *CYP52D4A* | TCATTGAATGCTTCCAGGAACCTCGCCACATCCATC GAGAACCGG (SEQ ID NO: 70) |
| 7581-97-A | F | *CYP52A5A* & *CYP52A5B* | GGATCCTAATACGACTCACTATAGGGAGGAAGAGG GCAGGGCTCAAGAG (SEQ ID NO: 59) |
| 7581-97-B | B | *CYP52A5A* & *CYP52A5B* | TCCATGTGAAGATCCCATCACGAGTGTGCCTCTTGC CCAAAG (SEQ ID NO: 60) |
| 4P-2/T7 | F | *CYP52A8A* | GGATCCTAATACGACTCACTATAGGGAGGCTTGAA GGCCGTGTTGAACG (SEQ ID NO: 71) |
| 4M-3/4M-1 | B | *CYP52A8A* | CAGGATTTGTCTGAGTTGCCGCCTGATCAAGATAG GATCCTTGCCG (SEQ ID NO: 72) |
| 3737-26-D | F | *CPRA* | GGATCCTAATACGACTCACTATAGGGAGGGGTTTG CTGAATACGCTGAAGGTGATG (SEQ ID NO: 73) |
| 3737-26-C | B | *CPRA* | TGGAGCTGAACAACTCTCTCGTCTCGGGTGGTCGA ATGGACCCTTGGTCAAG (SEQ ID NO: 74) |
| 3737-133C | F | *CPRA* & *CPRB* | GGATCCTAATACGACTCACTATAGGGAGGTTCCTC AACACGGACAGCGG (SEQ ID NO: 75) |
| 3737-133D | B | *CPRA* & *CPRB* | AGTCAACCAGGTGTGGAACTCGTCGGTGGCAACAA TGAAAAACACCAAG (SEQ ID NO: 76) |
| 3737-52-C | F | *POX4A* & *POX4B* | GGATCCTAATACGACTCACTATAGGGAGGCCATTG CCTTGAGATACGCCATTGGTAG (SEQ ID NO: 77) |
| 3737-52-D | B | *POX4A* & *POX4B* | AGCCTTGGTGTCGTTCTTTTCAACGGAAGGTGGTCT CGATGGTGTGTTCAACC (SEQ ID NO: 78) |
| 3737-53-C | F | *POX5A* | GGATCCTAATACGACTCACTATAGGGAGGTTGGGT TTGTTTGTTTCCTGTGTCCG (SEQ ID NO: 79) |
| 3737-53-D | B | *POX5A* | CCTTTGACCTTCAATCTGGCGTAGACGCAGCACCA CCGATCCACCACTTG (SEQ ID NO: 80) |

F=Forward B=Backword

The forward primer (SEQ ID NO: 59) contains the T7 promoter consensus sequence "GGATCCTAATACGA CTCACTATAGGG AGG" fused to the primer 7581-97-F sequence (SEQ ID NO: 47). The Reverse Primer (SEQ ID NO: 60) contains the sequence of primer 7581-97M (SEQ ID NO: 48) followed by the 20 bases of upstream sequence with a 18 base pair deletion between the two blocks of the *CYP52A5* sequence. The forward primer was used with the corresponding reverse primer to synthesize the competitor DNA template. The primer pairs were combined in a standard *Taq* Gold polymerase PCR reaction according to the manufacturer's recommended conditions (Perkin-Elmer/Applied Biosystems, Foster City, CA). The PCR reaction mix contained a final concentration of 250 nM each primer and 10 ng *C. tropicalis* chromosomal DNA for template. The reaction mixture was placed in a thermocycler for 25 to 35 cycles using the highest annealing temperature possible during the PCR reactions to assure a homogeneous PCR product (in this case 62°C). The PCR products were either gel purified or filtered purified to remove un-incorporated nucleotides and primers. The competitor template DNA was then quantified using the ($A_{260/280}$) method. Primers used in QC-RT-PCR experiments for the synthesis of various competitive DNA templates are listed in Table 7 (SEQ ID NOS: 61-80).

**C. Synthesis of the Competitor RNA**

Competitor template DNA was transcribed *In-Vitro* to make the competitor RNA using the Megascript T7 kit from Ambion Biosciences (Ambion Inc., Austin, Texas). 250 nanograms (ng) of competitor DNA template and the *in-vitro* transcription reagents are mixed according to the directions provided by the manufacturer. The reaction mixture was incubated for 4 hours at 37°C. The resulting RNA preparations were then checked by gel electrophoresis for the conditions giving the highest yields and quality of competitor RNA. This often required optimization according to the manufacturer's specifications. The DNA template was then removed using DNase I as described in the Ambion kit. The RNA competitor was then quantified by the ($A_{260/280}$) method. Serial dilution's of the RNA (1 ng/μl to 1 femtogram (fg)/μl) were made for use in the QC-RT-PCR reactions and the original stocks stored at -70°C.

### D. QC-RT-PCR Reactions

QC-RT-PCR reactions were performed using rTth polymerase from Perkin-Elmer(Perkin-Elmer/Applied Biosystems, Foster City, CA) according to the manufacturer's recommended conditions. The reverse transcription reaction was performed in a 10 µl volume with a final concentrations of 200 µM for each dNTP, 1.25 units rTth polymerase, 1.0 mM $MnCl_2$, 1X of the 10X buffer supplied with the Enzyme from the manufacturer, 100 ng of total RNA isolated from a fermentor grown culture of *C. tropicalis* and 1.25 µM of the appropriate reverse primer. To quantitate *CYP52A5* expression in *C. tropicalis* an appropriate reverse primer was 7581-97M (SEQ ID NO: 48). Several reaction mixes were prepared for each RNA sample characterized. To quantitate *CYP52A5* expression a series of 8 to 12 of the previously described QC-RT-PCR reaction mixes were aliquoted to different reaction tubes. To each tube 1 µl of a serial dilution containing from 100 pg to100 fg *CYP52A5* competitor RNA per µl was added bringing the final reaction mixtures up to the final volume of 10 µl. The QC-RT-PCR reaction mixtures were mixed and incubated at 70°C for 15 min according to the manufacturer's recommended times for reverse transcription to occur. At the completion of the 15 minute incubation, the sample temperature was reduced to 4°C to stop the reaction and 40 µl of the PCR reaction mix added to the reaction to bring the total volume up to 50 µl. The PCR reaction mix consists of an aqueous solution containing 0.3125 µM of the forward primer 7581-97F (SEQ ID NO: 47), 3.125 mM $MgCl_2$ and 1X chelating buffer supplied with the enzyme from Perkin-Elmer. The reaction mixtures were placed in a thermocycler (Perkin-Elmer GeneAmp PCR System 2400, Perkin-Elmer/Applied Biosystems, Foster City, CA ) and the following PCR cycle performed: 94°C for 1 min. followed by 94°C for 10 seconds followed by 58°C for 40 seconds for 17 to 22 cycles. The PCR reaction was completed with a final incubation at 58°C for 2 min followed by 4°C. In some reactions where no detectable PCR products were produced the samples were returned the thermocycler for additional cycles, this process was repeated until enough PCR products were produced to quantify using HPLC. The number of cycles necessary to produce enough PCR product is a function of the amount of the target mRNA in the 100 ng of total cellular RNA. In cultures where the *CYP52A5* gene is highly expressed there is sufficient *CYP52A5* mRNA message present and less PCR cycles (≤17) are required to produce quantifiable amount of PCR product. The lower the concentrations of the target mRNA present the more PCR cycles are required to produce a detectable amount of product. These QC-RT-

PCR procedures were applied to all the target genes listed in Table 5 using the respective primers indicated therein.

### E. HPLC Quantification

Upon completion of the QC-RT-PCR reactions the samples were analyzed and quantitated by HPLC. Five to fifteen microliters of the QC-RT-PCR reaction mix was injected into a Waters Bio-Compatible 625 HPLC with an attached Waters 484 tunable detector. The detector was set to measure a wave length of 254 nm. The HPLC contained a Sarasep brand DNASep™ column (Sarasep, Inc., San Jose, CA) which was placed within the oven and the temperature set for 52 °C. The column was installed according to the manufacturer's recommendation of having 30 cm. of heated PEEK tubing installed between the injector and the column. The system was configured with a Sarasep brand Guard column positioned before the injector. In addition, there was a 0.22 μm filter disk just before the column, within the oven. Two Buffers were used to create an elution gradient to resolve and quantitate the PCR products from the QC-RT-PCR reactions. Buffer-A consists of 0.1 M tri-ethyl ammonium acetate (TEAA) and 5% acetonitrile (volume to volume). Buffer-B consists of 0.1 M TEAA and 25% acetonitrile (volume to volume). The QC-RT-PCR samples were injected into the HPLC and the linear gradient of 75% buffer-A/ 25% buffer-B to 45% buffer-A/ 55% B was run over 6 min at a flow rate of 0.85 ml per minute. The QC-RT-PCR product of the competitor RNA being 18 base pairs smaller is eluted from the HPLC column before the QC-RT-PCR product from the *CYP52A5* mRNA(U). The amount of the QC-RT-PCR products are plotted and quantitated with an attached Waters Corporation 745 data module. The log ratios of the amount of *CYP52A5* mRNA QC-RT-PCR product (U) to competitor QC-RT-PCR product (C), as measured by peak areas, was plotted and the amount of competitor RNA required to equal the amount of *CYP52A5* mRNA product determined. In the case of each of the target genes listed in Table 5, the competitor RNA contained fewer base pairs as compared to the native target mRNA and eluted before the native mRNA in a manner similar to that demonstrated by *CYP52A5*. HPLC quantification of the genes was conducted as above.

## EXAMPLE 12

### Evaluation of New Strains in Shake Flasks

The *CYP* and *CPR* amplified strains such as strains HDC10, HDC15, HDC20 and HDC23 (Table 1) and H5343 were evaluated for diacid production in shake flasks. A single colony for each strain was transferred from a YPD agar plate into 5 ml of YPD broth and grown overnight at 30°C, 250 rpm. An inoculum was then transferred into 50 ml of DCA2 medium (Chart) and grown for 24 h at 30°C, 300 rpm. The cells were centrifuged at 5000 rpm for 5 min and resuspended in 50 ml of DCA3 medium (Chart) and grown for 24 h at 30°C, 300 rpm. 3% oleic acid w/v was added after 24 h growth in DCA3 medium and the cultures were allowed to bioconvert oleic acid for 48 h. Samples were harvested and the diacid and monoacid concentrations were analyzed as per the scheme given in Figure 35. Each strain was tested in duplicate and the results shown in Table 8 represent the average value from two flasks.

**Table 8.** Bioconversion of oleic acid by different recombinant strains of *Candida tropicalis*

| Strain | Conversion to Oleic diacid (%) | Specific Conversion (g diacid/g biomass |
|---|---|---|
| H5343 | 41.9 | 0.53 |
| HDC 10-2 | 50.5 | 0.85 |
| HDC 15 | 54.4 | 0.85 |
| HDC 20-1 | 45.1 | 0.72 |
| HDC 20-2 | 45.3 | 0.58 |
| HDC 23-2 | 55.2 | 0.84 |
| HDC 23-3 | 58.8 | 0.89 |

## EXAMPLE 13

### Cloning and Characterization of *C. tropicalis* 20336 Cytochrome P450 Monooxygenase (*CYP*) and Cytochrome P450 NADPH Oxidoreductase (*CPR*) Genes

To clone *CYP* and *CPR* genes several different strategies were employed. Available CYP amino acid sequences were aligned and regions of similarity were observed (Figure 4). These regions corresponded to described conserved regions seen in other cytochrome P450 families (Goeptar et al., *supra* and Kalb et al. *supra*). Proteins from eight eukaryotic

cytochrome P450 families share a segmented region of sequence similarity. One region corresponded to the HR2 domain containing the invariant cysteine residue near the carboxyl terminus which is required for heme binding while the other region corresponded to the central region of the I helix thought to be involved in substrate recognition (Figure 4). Degenerate oligonucleotide primers corresponding to these highly conserved regions of the *CYP52* gene family present in *Candida maltosa* and *Candida tropicalis* ATCC 750 were designed and used to amplify DNA fragments of *CYP* genes from *C. tropicalis* 20336 genomic DNA. These discrete PCR fragments were then used as probes to isolate full-length *CYP* genes from the *C. tropicalis* 20336 genomic libraries. In a few instances oligonucleotide primers corresponding to highly conserved regions were directly used as probes to isolate full-length *CYP* genes from genomic libraries. In the case of *CPR* a heterologous probe based upon the known DNA sequence for the *CPR* gene from *C. tropicalis* 750 was used to isolate the *C. tropicalis* 20336 *CPR* gene.

**A. Cloning of the *CPR* Gene from *C. tropicalis* 20336**

**1) Cloning of the *CPRA* Allele**

Approximately 25,000 phage particles from the first genomic library of *C. tropicalis* 20336 were screened with a 1.9 kb *Bam*HI-*Nde*I fragment from plasmid pCU3RED (See Picattagio et al., Bio/Technology 10:894-898 (1992), incorporated herein by reference) containing most of the *C. tropicalis* 750 *CPR* gene. Five clones that hybridized to the probe were isolated and the plasmid DNA from these lambda clones was rescued and characterized by restriction enzyme analysis. The restriction enzyme analysis suggested that all five clones were identical but it was not clear that a complete *CPR* gene was present.

PCR analysis was used to determine if a complete *CPR* gene was present in any of the five clones. Degenerate primers were prepared for highly conserved regions of known *CPR* genes (See Sutter et al., *J. Biol. Chem.* 265:16428-16436 (1990), incorporated herein by reference) ( Figure 4). Two Primers were synthesized for the FMN binding region (FMN1, SEQ ID NO: 16 and FMN2, SEQ ID NO: 17). One primer was synthesized for the FAD binding region (FAD, SEQ ID NO: 18), and one primer for the NADPH binding region (NADPH, SEQ ID NO: 19) (Table 4). These four primers were used in PCR amplification experiments using as a template plasmid DNA isolated from four of the five clones described above. The FMN (SEQ

ID NOS: 16 and 17) and FAD (SEQ ID NO: 18) primers served as forward primers and the NADPH primer (SEQ ID NO: 19) as the reverse primer in the PCR reactions. When different combinations of forward and reverse primers were used, no PCR products were obtained from any of the plasmids. However, all primer combinations amplified expected size products with a plasmid containing the *C. tropicalis* 750 *CPR* gene (positive control). The most likely reason for the failure of the primer pairs to amplify a product, was that all four of clones contained a truncated *CPR* gene. One of the four clones (pHKM1) was sequenced using the Triplex 5' (SEQ ID NO: 30) and the Triplex 3' (SEQ ID NO: 31) primers (Table 4) which flank the insert and the multiple cloning site on the cloning vector, and with the degenerate primer based upon the NADPH binding site described above. The NADPH primer (SEQ ID NO: 19) failed to yield any sequence data and this is consistent with the PCR analysis. Sequences obtained with Triplex primers were compared with *C. tropicalis* 750 *CPR* sequence using the MacVector™ program (Oxford Molecular Group, Campbell, CA). Sequence obtained with the Triplex 3' primer (SEQ ID NO: 31) showed similarity to an internal sequence of the *C. tropicalis* 750 *CPR* gene confirming that pHKM1 contained a truncated version of a 20336 *CPR* gene. pHKM1 had a 3.8 kb insert which included a 1.2 kb coding region of the *CPR* gene accompanied by 2.5 kb of upstream DNA (Figure 5). Approximately 0.85 kb of the 20336 *CPR* gene encoding the C-terminal portion of the *CPR* protein is missing from this clone.

Since the first Clontech library yielded only a truncated *CPR* gene, the second library prepared by Clontech was screened to isolate a full-length *CPR* gene. Three putative *CPR* clones were obtained. The three clones, having inserts in the range of 5-7 kb, were designated pHKM2, pHKM3 and pHKM4. All three were characterized by PCR using the degenerate primers described above. Both pHKM2 and pHKM4 gave PCR products with two sets of internal primers. pHKM3 gave a PCR product only with the FAD (SEQ ID NO: 18) and NADPH (SEQ ID NO: 19) primers suggesting that this clone likely contained a truncated *CPR* gene. All three plasmids were partially sequenced using the two Triplex primers and a third primer whose sequence was selected from the DNA sequence near the truncated end of the *CPR* gene present in pHKM1. This analysis confirmed that both pHKM2 & 4 have sequences that overlap pHKM1 and that both contained the 3' region of *CPR* gene that is missing from pHKM1. Portions of inserts from pHKM1 and pHKM4 were sequenced and a full-length *CPR* gene was identified. Based on the DNA sequence and PCR analysis, it was concluded that

pHKM1 contained the putative promoter region and 1.2 kb of sequence encoding a portion (5' end) of a *CPR* gene. pHKM4 had 1.1 kb of DNA that overlapped pHKM1 and contained the remainder (3' end) of a *CPR* gene along with a downstream untranslated region (Figure 6). Together these two plasmids contained a complete *CPRA* gene with an upstream promoter region. *CPRA* is 4206 nucleotides in length (SEQ ID NO: 81) and includes a regulatory region and a protein coding region (defined by nucleotides 1006-3042) which is 2037 base pairs in length and codes for a putative protein of 679 amino acids (SEQ ID NO: 83) (Figures 13 and 14). In Figure 13, the asterisks denote conserved nucleotides between *CPRA* and *CPRB*, bold denotes protein coding nucleotides, and the start and stop codons are underlined. The *CPRA* protein, when analyzed by the protein alignment program of the GeneWorks™ software package (Oxford Molecular Group, Campbell, CA), showed extensive homology to *CPR* proteins from *C. tropicalis* 750 and *C. maltosa*.

2) **Cloning of the *CPRB* Allele**

To clone the second *CPRB* allele, the third genomic library, prepared by Henkel, was screened using DNA fragments from pHKM1 and pHKM4 as probes. Five clones were obtained and these were sequenced with the three internal primers used to sequence *CPRA*. These primers were designated PRK1.F3 (SEQ ID NO: 20) , PRK1.F5 (SEQ ID NO: 21) and PRK4.R20 (SEQ ID NO: 22) (Table 4). and the two outside primers (M13 -20 and T3 [Stratagene]) for the polylinker region present in the pBK-CMV cloning vector. Sequence analysis suggested that four of these clones, designated pHKM5 to 8, contained inserts which were identical to the *CPRA* allele isolated earlier. All four seemed to contain a full length *CPR* gene. The fifth clone was very similar to the *CPRA* allele, especially in the open reading frame region where the identity was very high. However, there were significant differences in the 5' and 3' untranslated regions. This suggested that the fifth clone was the allele to *CPRA*. The plasmid was designated pHKM9 (Figure 7) and a 4.14 kb region of this plasmid was sequenced and the analysis of this sequence confirmed the presence of the *CPRB* allele (SEQ ID NO: 82), which includes a regulatory region and a protein coding region (defined by nucleotides 1033-3069) (Figure 13). The amino acid sequence of the *CPRB* protein is set forth in SEQ ID NO: 84 (Figure 14).

### B. Cloning of *C. tropicalis* 20336 (*CYP*) Genes

#### 1) Cloning of *CYP52A2A*, *CYP52A3A & 3B* and *CYP52A5A & 5B*

Clones carrying *CYP52A2A, A3A, A3B, A5A* and *A5B* genes were isolated from the first and second Clontech genomic libraries using an oligonucleotide probe (HemeB1, SEQ ID NO: 27) whose sequence was based upon the amino acid sequence for the highly conserved heme binding region present throughout the *CYP52* family. The first and second libraries were converted to the plasmid form and screened by colony hybridizations using the HemeB1 probe (SEQ ID NO: 27) (Table 4). Several potential clones were isolated and the plasmid DNA was isolated from these clones and sequenced using the HemeB1 oligonucleotide (SEQ ID NO: 27) as a primer. This approach succeeded in identifying five *CYP52* genes. Three of the *CYP* genes appeared unique, while the remaining two were classified as alleles. Based upon an arbitrary choice of homology to *CYP52* genes from *Candida maltosa*, these five genes and corresponding plasmids were designated *CYP52A2A* (pPA15 [Figure 26]), *CYP52A3A* (pPA57 [Figure 29]), *CYP52A3B* (pPA62 [Figure 30]), *CYP52A5A* (pPAL3 [Figure 31]) and *CYP52A5B* (pPA5 [Figure 32]). The complete DNA sequence including regulatory and protein coding regions of these five genes was obtained and confirmed that all five were *CYP52* genes (Figure 15). In Figure 15, the asterisks denote conserved nucleotides among the *CYP* genes. Bold indicates the protein coding nucleotides of the *CYP* genes, and the start and stop codons are underlined. The *CYP52A2A* gene as represented by SEQ ID NO: 86 has a protein coding region defined by nucleotides 1199-2767 and the encoded protein has an amino acid sequence as set forth in SEQ ID NO: 96. The *CYP52A3A* gene as represented by SEQ ID NO: 88 has a protein encoding region defined by nucleotides 1126-2748 and the encoded protein has an amino acid sequence as set forth in SEQ ID NO: 98. The *CYP52A3B* gene as represented by SEQ ID NO: 89 has a protein coding defined by nucleotides 913-2535 and the encoded protein has an amino acid sequence as set forth in SEQ ID NO: 99. The *CYP52A5A* gene as represented by SEQ ID NO: 90 has a protein coding region defined by nucleotides 1103-2656 and the encoded protein has an amino acid sequence as set forth in SEQ ID NO: 100. The *CYP52A5B* gene as represented by SEQ ID NO: 91 has a protein coding region defined by nucleotides 1142-2695 and the encoded protein has an amino acid sequence as set forth in SEQ ID NO: 101.

2) **Cloning of *CYP52A1A* and *CYP52A8A***

*CYP52A1A* and *CYP52A8A* genes were isolated from the third genomic library using PCR fragments as probes. The PCR fragment probe for *CYP52A1* was generated after PCR amplification of 20336 genomic DNA with oligonucleotide primers that were designed to amplify a region from the Helix I region to the HR2 region using all available *CYP52* genes from National Center for Biotechnology Information. Degenerate forward primers UCup1 (SEQ ID NO: 23) and UCup2 (SEQ ID NO: 24) were designed based upon an amino acid sequence (-RDTTAG-) from the Helix I region (Table 4). Degenerate primers UCdown1 (SEQ ID NO: 25) and UCdown2 (SEQ ID NO: 26) were designed based upon an amino acid sequence (-GQQFAL-) from the HR2 region (Table 4). For the reverse primers, the DNA sequence represents the reverse complement of the corresponding amino acid sequence. These primers were used in pairwise combinations in a PCR reaction with Stoffel *Taq* DNA polymerase (Perkin-Elmer Cetus, Foster City, CA) according to the manufacturer's recommended procedure. A PCR product of approximately 450 bp was obtained. This product was purified from agarose gel using Gene-clean™ (Bio 101, LaJolla, CA) and ligated to the pTAG™ vector (Figure 17) (R&D systems, Minneapolis, MN) according to the recommendations of the manufacturer. No treatment was necessary to clone into pTAG because it employs the use of the TA cloning technique. Plasmids from several transformants were isolated and their inserts were characterized. One plasmid contained the PCR clone intact. The DNA sequence of the PCR fragment (designated 44*CYP*3, SEQ ID NO: 107) shared homology with the DNA sequences for the *CYP52A1* gene of *C. maltosa* and the *CYP52A3* gene of *C. tropicalis 750.* This fragment was used as a probe in isolating the *C. tropicalis* 20336 *CYP52A1* homolog. The third genomic library was screened using the 44*CYP*3 PCR probe (SEQ ID NO: 107) and a clone (pHKM11) that contained a full-length *CYP52* gene was obtained (Figure 8). The clone contained a gene having regulatory and protein coding regions. An open reading frame of 1572 nucleotides encoded a *CYP52* protein of 523 amino acids (Figures 15 and 16 ). This *CYP52* gene was designated *CYP52A1A* (SEQ ID NO: 85) since its putative amino acid sequence (SEQ ID NO: 95) was most similar to the *CYP52A1* protein of *C. maltosa*. The protein coding region of the *CYP52A1A* gene is defined by nucleotides 1177-2748 of SEQ ID NO: 85.

A similar approach was taken to clone *CYP52A8A*. A PCR fragment probe for *CYP52A8* was generated using primers for highly conserved sequences of *CYP52A3*, *CYP52A2*

and *CYP52A5* genes of *C. tropicalis 750.* The reverse primer (primer 2,3,5,M) (SEQ ID NO: 29) was designed based on the highly conserved heme binding region (Table 4). The design of the forward primer (primer 2,3,5,P) (SEQ ID NO: 28) was based upon a sequence conserved near the N-terminus of the *CYP52A3*, *CYP52A2* and *CYP52A5* genes from *C. tropicalis 750* (Table 4). Amplification of 20336 genomic DNA with these two primers gave a mixed PCR product. One amplified PCR fragment was 1006 bp long (designated DCA1002). The DNA sequence for this fragment was determined and was found to have 85% identity to the DNA sequence for the *CYP52D4* gene of *C. tropicalis* 750. When this PCR product was used to screen the third genomic library one clone (pHKM12) was identified that contained a full-length *CYP52* gene along with 5' and 3' flanking sequences (Figure 9). The *CYP52* gene included regulatory and protein coding regions with an open reading frame of 1539 nucleotides long which encoded a putative CYP52 protein of 512 amino acids (Figures 15 and 16 ). This gene was designated as *CYP52A8A* (SEQ ID NO: 92) since its amino acid sequence (SEQ ID NO: 102) was most similar to the *CYP52A8* protein of *C. maltosa.* The protein coding region of the *CYP52A8A* gene is defined by nucleotides 464-2002 of SEQ ID NO: 92. The amino acid sequence of the *CYP52A8A* protein is set forth in SEQ ID NO: 102.

### 3) **Cloning of *CYP52D4A***

The screening of the second genomic library with the HemeB1 (SEQ ID NO: 27) primer (Table 4) yielded a clone carrying a plasmid (pPA18) that contained a truncated gene having homology with the *CYP52D4* gene of *C. maltosa* (Figure 33). A 1.3 to 1.5-kb *Eco*RI-*Sst*I fragment from pPA18 containing part of the truncated *CYP* gene was isolated and used as a probe to screen the third genomic library for a full length *CYP52* gene. One clone (pHKM13) was isolated and found to contain a full-length *CYP* gene with extensive 5' and 3' flanking sequences (Figure 10). This gene has been designated as *CYP52D4A* (SEQ ID NO: 94) and the complete DNA including regulatory and protein coding regions (coding region defined by nucleotides 767-2266) and putative amino acid sequence (SEQ ID NO: 104) of this gene is shown in Figures 15 and 16. *CYP52D4A* (SEQ ID NO: 94) shares the greatest homology with the *CYP52D4* gene of *C. maltosa.*

#### 4) **Cloning of *CYP52A2B* and *CYP52A8B***

A mixed probe containing *CYP52A1A*, *A2A*, *A3A*, *D4A*, *A5A* and *A8A* genes was used to screen the third genomic library and several putative positive clones were identified. Seven of these were sequenced with the degenerate primers Cyp52a (SEQ ID NO: 32), Cyp52b (SEQ ID NO: 33), Cyp52c (SEQ ID NO: 34) and Cyp52d (SEQ ID NO: 35) shown in Table 4. These primers were designed from highly conserved regions of the four *CYP52* subfamilies, namely *CYP52A*, *B*, *C* & *D*. Sequences from two clones, pHKM14 and pHKM15 (Figures 11 and 12), shared considerable homology with DNA sequence of the *C. tropicalis* 20336 *CYP52A2* and *CYP52A8* genes, respectively. The complete DNA (SEQ ID NO: 87) including regulatory and protein coding regions (coding region defined by nucleotides 1072-2640) and putative amino acid sequence (SEQ ID NO: 97) of the *CYP52* gene present in pHKM14 suggested that it is *CYP52A2B* (Figures 15 and 16). The complete DNA (SEQ ID NO: 93) including regulatory and protein coding regions (coding region defined by nucleotides 1017-2555) and putative amino acid sequence (SEQ ID NO: 103) of the *CYP52* gene present in pHKM15 suggested that it is *CYP52A8B* (Figures 15 and 16).

## <u>EXAMPLE 14</u>

### Identification of *CYP* and *CPR* Genes Induced by Selected Fatty Acid and Alkane Substrates

Genes whose transcription is turned on by the presence of selected fatty acid or alkane substrates have been identified using the QC-RT-PCR assay. This assay was used to measure (*CYP*) and (*CPR*) gene expression in fermentor grown cultures *C. tropicalis* ATCC 20962. This method involves the isolation of total cellular RNA from cultures of *C. tropicalis* and the quantification of a specific mRNA within that sample through the design and use of sequence specific QC-RT-PCR primers and an RNA competitor. Quantification is achieved through the use of known concentrations of highly homologous competitor RNA in the QC-RT-PCR reactions. The resulting QC-RT-PCR amplified cDNA's are separated and quantitated through the use of ion pairing reverse phase HPLC. This assay was used to characterize the expression of *CYP52* genes of *C. tropicalis* ATCC 20962 in response to various fatty acid and alkane substrates. Genes which were induced were identified by the calculation of their mRNA concentration at various times before and after induction. Figure 18 provides an example of

how the concentration of mRNA for *CYP52A5* can be calculated using the QC-RT-PCR assay. The log ratio of unknown (U) to competitor product (C) is plotted versus the concentration of competitor RNA present in the QC-RT-PCR reactions. The concentration of competitor which results in a log ratio of U/C of zero, represents the point where the unknown messenger RNA concentration is equal to the concentration of the competitor. Figure 18 allows for the calculation of the amount of *CYP52A5* message present in 100 ng of total RNA isolated from cell samples taken at 0, 1, and 2 hours after the addition of Emersol® 267 in a fermentor run. From this analysis, it is possible to determine the concentration of the *CYP52A5* mRNA present in 100 ng of total cellular RNA. In the plot contained in Figure 18 it takes 0.46 pg of competitor to equal the number of mRNA's of *CYP52A5* in 100 ng of RNA isolated from cells just prior (time 0) to the addition of the substrate, Emersol® 267. In cell samples taken at one and two hours after the addition of Emersol® 267 it takes 5.5 and 8.5 pg of competitor RNA, respectively. This result demonstrates that *CYP52A5* (SEQ ID NOS: 90 and 91) is induced more than 18 fold within two hours after the addition of Emersol® 267. This type of analysis was used to demonstrate that *CYP52A5* (SEQ ID NO: 90 and 91) is induced by Emersol® 267. Figure 19 shows the relative amounts of *CYP52A5* (SEQ ID NOS: 90 and 91) expression in fermentor runs with and without Emersol® 267 as a substrate. The differences in the *CYP52A5* (SEQ. ID NOS: 90 and 91) expression patterns are due to the addition of Emersol® 267 to the fermentation medium.

This analysis clearly demonstrates that expression of *CYP52A5* (SEQ ID NOS: 90 and 91) in *C. tropicalis* 20962 is inducible by the addition of Emersol® 267 to the growth medium. This analysis was performed to characterize the expression of *CYP52A2A* (SEQ ID NO: 86), *CYP52A3AB* (SEQ ID NOS: 88 and 89), *CYP52A8A* (SEQ ID NO: 92), *CYP52A1A* (SEQ ID NO: 85), *CYP52D4A* (SEQ ID NO: 94) and *CPRB* (SEQ ID NO: 82) in response to the presence of Emersol® 267 in the fermentation medium (Figure 20). The results of these analysis' indicate, that like the *CYP52A5* gene (SEQ ID NOS: 90 and 91) of *C. tropicalis* 20962, the *CYP52A2A* gene (SEQ ID NO: 86) is inducible by Emersol® 267. A small induction is observed for *CYP52A1A* (SEQ ID NO: 85) and *CYP52A8A* (SEQ ID NO: 92). In contrast, any induction for *CYP52D4A* (SEQ ID NO: 94), *CYP52A3A* (SEQ ID NO: 88), *CYP52A3B* (SEQ ID NO: 89) is below the level of detection of the assay. *CPRB* (SEQ ID NO: 82) is moderately induced by Emersol® 267, four to five fold. The results of these analysis are summarized in

Figure 20. Figure 34 provides an example of selective induction of *CYP52A* genes. When pure fatty acid or alkanes are spiked into a fermentor containing *C. tropicalis* 20962 or a derivative thereof, the transcriptional activation of *CYP52A* genes was detected using the QC-RT-PCR assay. Figure 34 shows that pure oleic acid (C18:1) strongly induces *CYP52A2A* (SEQ ID NO: 86) while inducing *CYP52A5* (SEQ ID NOS: 90 and 91). In the same fermentor addition of pure alkane (tridecane) shows strong induction of both *CYP52A2A* (SEQ ID NO: 86) and *CYP52A1A* (SEQ ID NO: 85). However, tridecane did not induce *CYP52A5* (SEQ ID NOS: 90 and 91). In a separate fermentation using ATCC 20962, containing pure octadecane as the substrate, induction of *CYP52A2A*, *CYP52A5A* and *CYP52A1A* is detected (see Figure 36). The foregoing demonstrates selective induction of particular *CYP* genes by specific substrates, thus providing techniques for selective metabolic engineering of cell strains. For example, if tridecane modification is desired, organisms engineered for high levels of *CYP52A2A* (SEQ ID NO: 86) and *CYP52A1A* (SEQ ID NO: 85) activity are indicated. If oleic acid modification is desired, organisms engineered for high levels of *CYP52A2A* (SEQ ID NO: 86) activity are indicated.

## EXAMPLE 15

### Integration of Selected *CYP* and *CPR* Genes into the Genome of *Candida tropicalis*

In order to integrate selected genes into the chromosome of *C. tropicalis* 20336 or its descendants, there has to be a target DNA sequence, which may or may not be an intact gene, into which the genes can be inserted. There must also be a method to select for the integration event. In some cases the target DNA sequence and the selectable marker are the same and, if so, then there must also be a method to regain use of the target gene as a selectable marker following the integration event. In *C. tropicalis* and its descendants, one gene which fits these criteria is *URA3A*, encoding orotidine-5'-phosphate decarboxylase. Using it as a target for integration, *ura⁻* variants of *C. tropicalis* can be transformed in such a way as to regenerate a $URA^+$ genotype via homologous recombination (Figure 21). Depending upon the design of the integration vector, one or more genes can be integrated into the genome at the same time. Using a split *URA3A* gene oriented as shown in Figure 22, homologous integration would yield at least one copy of the gene(s) of interest which are inserted between the split portions of the *URA3A* gene. Moreover, because of the high sequence similarity between *URA3A* and *URA3B* genes, integration of the

construct can occur at both the *URA3A* and *URA3B* loci. Subsequently, an oligonucleotide designed with a deletion in a portion of the *URA* gene based on the identical sequence across both the *URA3A* and *URA3B* genes, can be utilized to yield *C. tropicalis* transformants which are once again *ura⁻* but which still carry one or more newly integrated genes of choice (Figure 21). *ura⁻* variants of *C. tropicalis* can also be isolated via other methods such as classical mutagenesis or by spontaneous mutation. Using well established protocols, selection of *ura⁻* strains can be facilitated by the use of 5-fluoroorotic acid (5-FOA) as described, e.g., in Boeke et al., *Mol. Gen. Genet.* 197:345-346, (1984), incorporated herein by reference. The utility of this approach for the manipulation of *C. tropicalis* has been well documented as described, e.g., in Picataggio et al., *Mol. and Cell. Biol.* 11:4333-4339 (1991); Rohrer et al., *Appl. Microbiol. Biotechnol.* 36:650-654 (1992); Picataggio et al., *Bio/Technology* 10:894-898 (1992); U.S. Patent No. 5,648,247; U.S. Patent No. 5,620,878; U.S. Patent No. 5,204,252; U.S. Patent No. 5,254,466, all of which are incorporated herein by reference.

**A. Construction of a URA Integration Vector, pURAin.**

Primers were designed and synthesized based on the 1712 bp sequence of the *URA3A* gene of *C. tropicalis* 20336 (see Figure 23). The nucleotide sequence of the *URA3A* gene of *C. tropicalis* 20336 is set forth in SEQ ID NO: 105 and the amino acid sequence of the encoded protein is set forth in SEQ ID NO: 106. *URA3A* Primer Set #1a (SEQ ID NO: 9) and #1b (SEQ ID NO: 10) (Table 4) was used in PCR with *C. tropicalis* 20336 genomic DNA to amplify *URA3A* sequences between nucleotide 733 and 1688 as shown in Figure 23. The primers are designed to introduce unique 5' *Asc*I and 3' *Pac*I restriction sites into the resulting amplified *URA3A* fragment. *Asc*I and *Pac*I sites were chosen because these sites are not present within *CYP* or *CPR* genes identified to date. *URA3A* Primer Set #2 was used in PCR with *C. tropicalis* 20336 genomic DNA as a template, to amplify *URA3A* sequences between nucleotide 9 and 758 as shown in Figure 23. *URA3A* Primer set #2a (SEQ ID NO: 11) and #2b (SEQ ID NO: 12) (Table 4) was designed to introduce unique 5' *Pac*I and 3' *Pme*I restriction sites into the resulting amplified *URA3A* fragment. The *Pme*I site is also not present within *CYP* and *CPR* genes identified to date. PCR fragments of the *URA3A* gene were purified, restricted with *Asc*I, *Pac*I and *Pme*I restriction enzymes and ligated to a gel purified, QiaexII cleaned *Asc*I-*Pme*I digest of plasmid pNEB193 (Figure 25) purchased from New England Biolabs (Beverly, MA).

The ligation was performed with an equimolar number of DNA termini at 16 °C for 16 hr using T4 DNA ligase (New England Biolabs). Ligations were transformed into *E. coli* XL1-Blue cells (Stratagene, LaJolla, CA) according to manufacturers recommendations. White colonies were isolated, grown, plasmid DNA isolated and digested with *Asc*I-*Pme*I to confirm insertion of the modified *URA3A* into pNEB193. The resulting base integration vector was named pURAin (Figure 24).

### B. Amplification of *CYP52A2A*, *CYP52A3A*, *CYP52A5A* and *CPRB* from *C. tropicalis* 20336 Genomic DNA

The genes encoding *CYP52A2A*, (SEQ ID NO: 86) and *CYP52A3A* (SEQ ID NO: 88) from *C. tropicalis* 20336 were amplified from genomic clones (pPA15 and pPA57, respectively) (Figures 26 and 29) via PCR using primers (Primer *CYP* 2A#1, SEQ ID NO: 1 and Primer *CYP* 2A#2, SEQ ID NO: 2 for CYP52A2A) (Primer *CYP* 3A#1, SEQ ID NO: 3 and Primer *CYP* 3A#2, SEQ ID NO: 4 for CYP52A3A) to introduce *Pac*I cloning sites. These PCR primers were designed based upon the DNA sequence determined for *CYP52A2A* (SEQ ID NO: 86) (Figure 15). The Ampli*Taq* Gold PCR kit (Perkin Elmer Cetus, Foster City, CA) was used according to manufacturers specifications. The *CYP52A2A* PCR amplification product was 2,230 base pairs in length , yielding 496 bp of DNA upstream of the *CYP52A2A* start codon and 168 bp downstream of the stop codon for the *CYP52A2A* ORF. The *CYP52A3A* PCR amplification product was 2154 base pairs in length, yielding 437bp of DNA upstream of the *CYP52A3A* start codon and 97bp downstream of the stop codon for the *CYP52A3A ORF.* The *CYP52A3A* PCR amplification product was 2154 base pairs in length, yielding 437bp of DNA upstream of the *CYP52A3A* start codon and 97bp downsteam of the stop codon for the *CYP52A3A* ORF.

The gene encoding *CYP52A5A* (SEQ ID NO: 90) from *C. tropicalis* 20336 was amplified from genomic DNA via PCR using primers (Primer *CYP* 5A#1, SEQ ID NO: 5 and Primer *CYP* 5A#2, SEQ ID NO: 6) to introduce *Pac*I cloning sites. These PCR primers were designed based upon the DNA sequence determined for *CYP52A5A* (SEQ ID NO: 90) . The Expand Hi-Fi *Taq* PCR kit (Boehringer Mannheim, Indianapolis, IN) was used according to manufacturers specifications. The *CYP52A5A* PCR amplification product was 3,298 base pairs in length.

The gene encoding *CPRB* (SEQ ID NO: 82) from *C. tropicalis* 20336 was amplified from genomic DNA via PCR using primers (*CPR* B#1, SEQ ID NO: 7 and *CPR* B#2, SEQ ID NO: 8) based upon the DNA sequence determined for *CPRB* (SEQ ID NO: 82) (Figure 13). These primers were designed to introduce unique *Pac*I cloning sites. The Expand Hi-Fi *Taq* PCR kit (Boehringer Mannheim, Indianapolis, IN) was used according to manufacturers specifications. The *CPRB* PCR product was 3266 bp in length, yielding 747 bp pf DNA upstream of the *CPRB* start codon and 493 bp downstream of the stop codon for the *CPRB* ORF. The resulting PCR products were isolated via agarose gel electrophoresis, purified using QiaexII and digested with *Pac*I. The PCR fragments were purified, desalted and concentrated using a Microcon 100 (Amicon, Beverly, MA).

The above described amplification procedures are applicable to the other genes listed in Table 5 using the respectively indicated primers.

### C. Cloning of *CYP* and *CPR* Genes into pURAin.

The next step was to clone the selected *CYP* and *CPR* genes into the pURAin integration vector. In a preferred aspect of the present invention, no foreign DNA other than that specifically provided by synthetic restriction site sequences are incorporated into the DNA which was cloned into the genome of *C. tropicalis,* i.e., with the exception of restriction site DNA only native *C. tropicalis* DNA sequences are incorporated into the genome. pURAin was digested with *Pac*I, Qiaex II cleaned, and dephosphorylated with Shrimp Alkaline Phosphatase (SAP) (United States Biochemical, Cleveland, OH) according the manufacturer's recommendations. Approximately 500 ng of *Pac*I linearized pURAin was dephosphorylated for 1 hr at 37°C using SAP at a concentration of 0.2 Units of enzyme per 1 pmol of DNA termini. The reaction was stopped by heat inactivation at 65°C for 20 min.

The *CYP52A2A PacI* fragment derived using the primer shown in Table 4 was ligated to plasmid pURAin which had also been digested with *Pac*I. *Pac*I digested pURAin was dephosphorylated, and ligated to the *CYP52A2A* ULTMA PCR product as described previously. The ligation mixture was transformed into *E. coli* XL1 Blue MRF' (Stratagene) and 2 resistant colonies were selected and screened for correct constructs which should contain vector sequence, the inverted *URA3A* gene, and the amplified *CYP52A2A* gene (SEQ ID NO: 86) of 20336. *AscI-PmeI* digestion identified one of the two constructs, plasmid pURA2in, as being correct (Figure

27). This plasmid was sequenced and compared to *CYP52A2A* (SEQ ID NO: 86) to confirm that PCR did not introduce DNA base changes that would result in an amino acid change.

Prior to its use, the *CPRB PacI* fragment derived using the primers shown in Table 4 was sequenced and compared to *CPRB* (SEQ ID NO: 82) to confirm that PCR did not introduce DNA base pair changes that would result in an amino acid change. Following confirmation, *CPRB* (SEQ ID NO: 82) was ligated to plasmid pURAin which had also been digested with *Pac*I. *Pac*I digested pURAin was dephosphorylated, and ligated to the *CPR* Expand Hi-Fi PCR product as described previously. The ligation mixture was transformed into *E. coli* XL1 Blue MRF' (Stratagene) and several resistant colonies were selected and screened for correct constructs which should contain vector sequence, the inverted *URA3A* gene, and the amplified *CPRB* gene (SEQ ID NO: 82) of 20336. AscI-*Pme*I digestion confirmed a successful construct, pURAREDBin.

In a manner similar to the above, each of the other *CYP* and *CPR* genes disclosed herein are cloned into pURAin. *Pac*I fragments of these genes, whose sequences are given in Figures 13 and 15, are derivable by methods known to those skilled in the art.

**1) Construction of Vectors Used to Generate HDC 20 and HDC 23**

A previously constructed integration vector containing *CPRB* (SEQ ID NO: 82), pURAREDBin, was chosen as the starting vector. This vector was partially digested with *Pac*I and the linearized fragment was gel-isolated. The active *Pac*I was destroyed by treatment with T4 DNA polymerase and the vector was re-ligated. Subsequent isolation and complete digestion of this new plasmid yielded a vector now containing only one active *Pac*I site. This fragment was gel-isolated, dephosphorylated and ligated to the *CYP52A2A PacI* fragment. Vectors that contain the *CYP52A2A* (SEQ ID NO: 86) and *CPRB* (SEQ ID NO: 82) genes oriented in the same direction, pURAin *CPR* 2A S, as well as opposite directions (5' ends connected), pURAin *CPR* 2A O, were generated.

**D. Confirmation of *CYP* Integration (Figure 21 for Integration Scheme) into the Genome of *C. tropicalis***

Based on the construct, pURA2in, used to transform H5343 *ura*⁻, a scheme to detect integration was devised. Genomic DNA from transformants was digested with *Dra* III

and *Spe* I which are enzymes that cut within the *URA3A*, and *URA3B* genes but not within the integrated *CYP52A2A* gene. Digestion of genomic DNA where an integration had occurred at the *URA3A* or *URA3B* loci would be expected to result in a 3.5 kb or a 3.3 kb fragment, respectively (Figure 28). Moreover, digestion of the same genomic DNA with *Pac*I would yield a 2.2 kb fragment characteristic for the integrated *CYP52A2A* gene (Figure 28). Southern hybridizations of these digests with fragments of the *CYP52A2A* gene were used to screen for these integration events. Intensity of the band signal from the Southern using *Pac*I digestion was used as a measure of the number of integration events, ((i.e. the more copies of the *CYP52A2A* gene (SEQ ID NO: 86) which are present, the stronger the hybridization signal)).

*C. tropicalis* H5343 transformed *URA* prototrophs were grown at 30°C, 170 rpm, in 10 ml SC-uracil media for preparation of genomic DNA. Genomic DNA was isolated by the method described previously. Genomic DNA was digested with *Spe*I and *Dra*III. A 0.95% agarose gel was used to prepare a Southern hybridization blot. The DNA from the gel was transferred to a MagnaCharge nylon filter membrane (MSI Technologies, Westboro, MA) according to the alkaline transfer method of Sambrook et al., *supra*. For the Southern hybridization, a 2.2 kb *CYP52A2A* DNA fragment was used as a hybridization probe. 300 ng of *CYP52A2A* DNA was labeled using a ECL Direct labeling and detection system (Amersham) and the Southern was processed according to the ECL kit specifications. The blot was processed in a volume of 30 ml of hybridization fluid corresponding to 0.125 ml/cm$^2$. Following a prehybridization at 42°C for 1 hr, 300 ng of *CYP52A2A* probe was added and the hybridization continued for 16 hr at 42°C. Following hybridization, the blots were washed two times for 20 min each at 42 °C in primary wash containing urea. Two 5 min secondary washes at RT were conducted, followed by detection according to directions. The blots were exposed for 16 hours (hr) as recommended.

Integration was confirmed by the detection of a *Spe*I-*Dra*III 3.5 kb fragment from the genomic DNA of the transformants but not with the *C. tropicalis* 20336 control. Subsequently, a *Pac*I digestion of the genomic DNA of the positive transformants, followed by a Southern hybridization using an *CYP52A2A* gene probe, confirmed integration by the detection of a 2.2 kb fragment. The resulting *CYP52A2A* integrated strain was named HDC1 (see Table 1).

In a manner similar to the above, each of the genes contained in the *Pac*I fragments which are described in Section 3c above were confirmed for integration into the genome of *C. tropicalis.*

Transformants generated by transformation with the vectors, pURAin *CPR* 2A S or pURAin *CPR* 2A O, were analyzed by Southern hybridization for integration of both the *CYP52A2A* (SEQ ID NO: 86) and *CPRB* (SEQ ID NO: 82) genes tandemly. Three strains were generated in which the *CYP52A2A* (SEQ ID NO: 86) and *CPRB* (SEQ ID NO: 82) genes integrated are in the opposite orientation (HDC 20-1, HDC 20-2 and HDC 20-3) and three were generated with the *CYP52A2A* (SEQ ID NO: 86) and *CPRB* (SEQ ID NO: 82) genes integrated in the same orientation (HDC 23-1, HDC 23-2 and HDC 23-3), Table 1.

**E. Confirmation of *CPRB* Integration into H5343 *ura*⁻**

Seven transformants were screened by colony PCR using *CPRB* primer #2 (SEQ ID NO: 8) and a *URA3A*- specific primer. In five of the transformants, successful integration was detected by the presence of a 3899 bp PCR product. This 3899 bp PCR product represents the *CPRB* gene adjacent to the *URA3A* gene in the genome of H5343 thereby confirming integration. The resulting *CPRB* integrated strains were named HDC10-1 and HDC10-2 (see Table 1).

**F. Strain Evaluation.**

As determined by quantitative PCR, when compared to parent H5343, HDC10-1 contained three additional copies of the reductase gene and HDC10-2 contained four additional copies of the reductase gene. Evaluations of HDC20-1, HDC20-2 and HDC20-3 based on Southern hybridization data indicates that HDC20-1 contained multiple integrations, i.e., 2 to 3 times that of HDC20-2 or HDC20-3. Evaluations of HDC23-1, HDC23-2, and HDC23-3 based on Southern hybridization data indicates that HDC23-3 contained multiple integrations, i.e., 2 to 3 times that of HDC23-1 or HDC23-2. The data in Table 8 indicates that the integration of components of the ω-hydroxylase complex have a positive effect on the improvement of *Candida tropicalis* ATCC 20962 as a biocatalyst. The results indicate that *CYP52A5A* (SEQ ID NO: 90) is an important gene for the conversion of oleic acid to diacid. Surprisingly, tandem integrations of *CYP* and *CPR* genes oriented in the opposite direction (HDC 20 strains) seem to

be less productive than tandem integrations oriented in the same direction (HDC 23 strains), Tables 1 and 8.

## CHART

Media Composition

LB Broth

| | |
|---|---|
| Bacto Tryptone | 10 g |
| Bacto Yeast Extract | 5 g |
| Sodium Chloride | 10 g |
| Distilled Water | 1,000 ml |

LB Agar

| | |
|---|---|
| Bacto Tryptone | 10 g |
| Bacto Yeast Extract | 5 g |
| Sodium Chloride | 10 g |
| Agar | 15 g |
| Distilled Water | 1,000 ml |

LB Top Agarose

| | |
|---|---|
| Bacto Tryptone | 10 g |
| Bacto Yeast Extract | 5 g |
| Sodium Chloride | 10 g |
| Agarose | 7 g |
| Distilled Water | 1,000 ml |

NZCYM Broth

| | |
|---|---|
| Bacto Casein Digest | 10 g |
| Bacto Casamino Acids | 1 g |
| Bacto Yeast Extract | 5 g |
| Sodium Chloride | 5 g |
| Magnesium Sulfate (anhydrous) | 0.98 g |
| Distilled Water | 1,000 ml |

NZCYM Agar

| | |
|---|---|
| Bacto Casein Digest | 10 g |
| Bacto Casamino Acids | 1 g |
| Bacto Yeast Extract | 5 g |
| Sodium Chloride | 5 g |
| Magnesium Sulfate (anhydrous) | 0.98 g |
| Agar | 15 g |
| Distilled Water | 1,000 ml |

NZCYM Top Agarose

| | |
|---|---|
| Bacto Casein Digest | 10 g |
| Bacto Casamino Acids | 1 g |
| Bacto Yeast Extract | 5 g |
| Sodium Chloride | 5 g |
| Magnesium Sulfate (anhydrous) | 0.98 g |
| Agarose | 7 g |
| Distilled Water | 1,000 ml |

YEPD Broth

| | |
|---|---|
| Bacto Yeast Extract | 10 g |
| Bacto Peptone | 20 g |
| Glucose | 20 g |
| Distilled Water | 1,000 ml |

YEPD Agar*

| | |
|---|---|
| Bacto Yeast Extract | 10 g |
| Bacto Peptone | 20 g |
| Glucose | 20 g |
| Agar | 20 g |
| Distilled Water | 1,000 ml |

SC - uracil*

| | |
|---|---|
| Bacto-yeast nitrogen base without amino acids | 6.7g |
| Glucose | 20g |
| Bacto-agar | 20g |
| Drop-out mix | 2g |
| Distilled water | 1,000ml |

| <u>DCA2 medium</u> | g/l |
|---|---|
| Peptone | 3.0 |
| Yeast Extract | 6.0 |
| Sodium Acetate | 3.0 |
| Yeast Nitrogen Base (Difco) | 6.7 |
| Glucose (anhydrous) | 50.0 |
| Potassium Phosphate (dibasic, trihydrate) | 7.2 |
| Potassium Phosphate (monobasic, anhydrous) | 9.3 |

| <u>DCA3 medium</u> | g/l |
|---|---|
| 0.3 M Phosphate buffer containing, pH 7.5 | |
| Glycerol | 50 |
| Yeast Nitrogen base (Difco) | 6.7 |

<u>Drop-out mix</u>

| | | | |
|---|---|---|---|
| Adenine | 0.5g | Alanine | 2g |
| Arginine | 2g | Asparagine | 2g |
| Aspartic acid | 2g | Cysteine | 2g |
| Glutamine | 2g | Glutamic acid | 2g |
| Glycine | 2g | Histidine | 2g |
| Inositol | 2g | Isoleucine | 2g |
| Leucine | 10g | Lysine | 2g |
| Methionine | 2g | para-Aminobenzoic acid | 0.2g |
| Phenylalanine | 2g | Proline | 2g |
| Serine | 2g | Threonine | 2g |
| Tryptophan | 2g | Tyrosine | 2g |
| Valine | 2g | | |

*See Kaiser et al., Methods in Yeast Genetics, Cold Spring Harbor Laboratory Press, USA (1994), incorporated herein by reference.

It will be understood that various modifications may be made to the embodiments and/or examples disclosed herein. Thus, the above description should not be construed as limiting, but merely as exemplifications of preferred embodiments. Those skilled in the art will envision other modifications within the scope and spirit of the claims appended hereto.

WHAT IS CLAIMED IS:

1. Isolated nucleic acid encoding a *CPRA* protein having the amino acid sequence set forth in SEQ ID NO: 83.

2. Isolated nucleic acid comprising a coding region defined by nucleotides 1006-3042 as set forth in SEQ ID NO: 81.

3. Isolated nucleic acid according to claim 2 comprising the nucleotide sequence as set forth in SEQ ID NO: 81.

4. Isolated protein comprising an amino acid sequence as set forth in SEQ ID NO: 83.

5. A vector comprising a nucleotide sequence encoding *CPRA* protein including an amino acid sequence as set forth in SEQ ID NO: 83.

6. A vector according to claim 5 wherein the nucleotide sequence is set forth in nucleotides 1006-3042 of SEQ ID NO: 81

7. A vector according to claim 5 wherein the vector is selected from the group consisting of plasmid, phagemid, phage and cosmid.

8. A host cell transfected or transformed with the nucleic acid of claim 1.

9. A host cell according to claim 8 wherein the host cell is a yeast cell.

10. A host cell according to claim 9 wherein the yeast cell is a *Candida sp.*

11. A host cell according to claim 10 wherein the *Candida sp.* is *Candida tropicalis*.

12. A host cell according to claim 11 wherein the *Candida tropicalis* is *Candida tropicalis* 20336.

13. A host cell according to claim 12 wherein the *Candida tropicalis* is H5343 ura-.

14. A method of producing a *CPRA* protein including an amino acid sequence as set forth in SEQ ID NO: 83 comprising:

a) transforming a suitable host cell with a DNA sequence that encodes the protein having the amino acid sequence as set forth in SEQ ID NO: 83; and

b) culturing the cell under conditions favoring the expression of the protein.

15. The method according to claim 14 wherein the step of culturing the cell comprises adding an organic substrate to media containing the cell.

16. Isolated nucleic acid encoding a *CPRB* protein having the amino acid sequence set forth in SEQ ID NO: 84.

17. Isolated nucleic acid comprising a coding region defined by nucleotides 1033-3069 as set forth in SEQ ID NO: 82.

18. Isolated nucleic acid according to claim 17 comprising the nucleotide sequence as set forth in SEQ ID NO: 82.

19. Isolated protein comprising an amino acid sequence as set forth in SEQ ID NO: 84.

20. A vector comprising a nucleotide sequence encoding *CPRB* protein including an amino acid sequence as set forth in SEQ ID NO: 84.

21. A vector according to claim 20 wherein the nucleotide sequence is set forth in nucleotides 1033-3069 of SEQ ID NO: 82.

22. A vector according to claim 20 wherein the vector is selected from the group consisting of plasmid, phagemid, phage and cosmid..

23. A host cell transfected or transformed with the nucleic acid of claim 16.

24. A host cell according to claim 23 wherein the host cell is a yeast cell.

25. A host cell according to claim 24 wherein the yeast cell is a *Candida sp.*

26. A host cell according to claim 25 wherein the *Candida sp.* is *Candida tropicalis.*

27. A host cell according to claim 26 wherein the *Candida tropicalis* is *Candida tropicalis* 20336.

28. A host cell according to claim 27 wherein the *Candida tropicalis* is H5343 ura-.

29. A method of producing a *CPRB* protein including an amino acid sequence as set forth in SEQ ID NO: 84 comprising:

a) transforming a suitable host cell with a DNA sequence that encodes the protein having the amino acid sequence as set forth in SEQ ID NO: 84; and

b) culturing the cell under conditions favoring the expression of the protein.

30. The method according to claim 29 wherein the step of culturing the cell comprises adding an organic substrate to media containing the cell.

31. Isolated nucleic acid encoding a *CYP52A1A* protein having the amino acid sequence set forth in SEQ ID NO: 95.

32. Isolated nucleic acid comprising a coding region defined by nucleotides 1177-2748 as set forth in SEQ ID NO: 85.

33. Isolated nucleic acid according to claim 32 comprising the nucleotide sequence as set forth in SEQ ID NO: 85.

34. Isolated protein comprising an amino acid sequence as set forth in SEQ ID NO: 95.

35. A vector comprising a nucleotide sequence encoding *CYP52A1A* protein including an amino acid sequence as set forth in SEQ ID NO: 95.

36. A vector according to claim 35 wherein the nucleotide sequence is set forth in nucleotides 1177-2748 of SEQ ID NO: 85.

37. A vector according to claim 35 wherein the vector is selected from the group consisting of plasmid, phagemid, phage and cosmid.

38. A host cell transfected or transformed with the nucleic acid of claim 31.

39. A host cell according to claim 38 wherein the host cell is a yeast cell.

40. A host cell according to claim 39 wherein the yeast cell is a *Candida sp.*

41. A host cell according to claim 40 wherein the *Candida sp.* is *Candida tropicalis.*

42. A host cell according to claim 41 wherein the *Candida tropicalis* is *Candida tropicalis* 20336.

43. A host cell according to claim 42 wherein the *Candida tropicalis* is H5343 ura-.

44. A method of producing a *CYP52A1A* protein including an amino acid sequence as set forth in SEQ ID NO: 95 comprising:

a) transforming a suitable host cell with a DNA sequence that encodes the protein having the amino acid sequence as set forth in SEQ ID NO: 95; and

b) culturing the cell under conditions favoring the expression of the protein.

45. The method according to claim 44 wherein the step of culturing the cell comprises adding an organic substrate to media containing the cell.

46. Isolated nucleic acid encoding a *CYP52A2A* protein having the amino acid sequence set forth in SEQ ID NO: 96.

47. Isolated nucleic acid comprising a coding region defined by nucleotides 1199-2767 as set forth in SEQ ID NO: 86.

48. Isolated nucleic acid according to claim 47 comprising the nucleotide sequence as set forth in SEQ ID NO: 86.

49. Isolated protein comprising an amino acid sequence as set forth in SEQ ID NO: 96.

50. A vector comprising a nucleotide sequence encoding *CYP52A2A* protein including an amino acid sequence as set forth in SEQ ID NO: 96.

51. A vector according to claim 50 wherein the nucleotide sequence is set forth in nucleotides 1199-2767 of SEQ ID NO: 86.

52. A vector according to claim 50 wherein the vector is selected from the group consisting of plasmid, phagemid, phage and cosmid.

53. A host cell transfected or transformed with the nucleic acid of claim 46.

54. A host cell according to claim 53 wherein the host cell is a yeast cell.

55. A host cell according to claim 54 wherein the yeast cell is a *Candida sp.*

56. A host cell according to claim 55 wherein the *Candida sp.* is *Candida tropicalis*.

57. A host cell according to claim 56 wherein the *Candida tropicalis* is *Candida tropicalis* 20336.

58. A host cell according to claim 57 wherein the *Candida tropicalis* is H5343 ura-.

59. A method of producing a *CYP52A2A* protein including an amino acid sequence as set forth in SEQ ID NO: 96 comprising:

a) transforming a suitable host cell with a DNA sequence that encodes the protein having the amino acid sequence as set forth in SEQ ID NO: 96; and

b) culturing the cell under conditions favoring the expression of the protein.

60. The method according to claim 59 wherein the step of culturing the cell comprises adding an organic substrate to media containing the cell.

61. Isolated nucleic acid encoding a *CYP52A2B* protein having the amino acid sequence set forth in SEQ ID NO: 97.

62. Isolated nucleic acid comprising a coding region defined by nucleotides 1072-2640 as set forth in SEQ ID NO: 87.

63. Isolated nucleic acid according to claim 62 comprising the nucleotide sequence as set forth in SEQ ID NO: 87.

64. Isolated protein comprising an amino acid sequence as set forth in SEQ ID NO: 97.

65. A vector comprising a nucleotide sequence encoding *CYP52A2B* protein including an amino acid sequence as set forth in SEQ ID NO: 97.

66. A vector according to claim 65 wherein the nucleotide sequence is set forth in nucleotides 1072-2640 of SEQ ID NO: 87.

67. A vector according to claim 65 wherein the vector is selected from the group consisting of plasmid, phagemid, phage and cosmid.

68. A host cell transfected or transformed with the nucleic acid of claim 61.

69. A host cell according to claim 68 wherein the host cell is a yeast cell.

70. A host cell according to claim 69 wherein the yeast cell is a *Candida sp.*

71. A host cell according to claim 70 wherein the *Candida sp.* is *Candida tropicalis.*

72. A host cell according to claim 71 wherein the *Candida tropicalis* is *Candida tropicalis* 20336.

73. A host cell according to claim 72 wherein the *Candida tropicalis* is H5343 ura-.

74. A method of producing a *CYP52A2B* protein including an amino acid sequence as set forth in SEQ ID NO: 97 comprising:

a) transforming a suitable host cell with a DNA sequence that encodes the protein having the amino acid sequence as set forth in SEQ ID NO: 97; and

b) culturing the cell under conditions favoring the expression of the protein.

75. The method according to claim 74 wherein the step of culturing the cell comprises adding an organic substrate to media containing the cell.

76. Isolated nucleic acid encoding a *CYP52A3A* protein having the amino acid sequence set forth in SEQ ID NO: 98.

77. Isolated nucleic acid comprising a coding region defined by nucleotides 1126-2748 as set forth in SEQ ID NO: 88.

78. Isolated nucleic acid according to claim 77 comprising the nucleotide sequence as set forth in SEQ ID NO: 88.

79. Isolated protein comprising an amino acid sequence as set forth in SEQ ID NO: 98.

80. A vector comprising a nucleotide sequence encoding *CYP52A3A* protein including an amino acid sequence as set forth in SEQ ID NO: 98.

81. A vector according to claim 80 wherein the nucleotide sequence is set forth in nucleotides 1126-2748 of SEQ ID NO: 88.

82. A vector according to claim 80 wherein the vector is selected from the group consisting of plasmid, phagemid, phage and cosmid.

83. A host cell transfected or transformed with the nucleic acid of claim 76.

84. A host cell according to claim 83 wherein the host cell is a yeast cell.

85. A host cell according to claim 84 wherein the yeast cell is a *Candida sp.*

86. A host cell according to claim 85 wherein the *Candida sp.* is *Candida tropicalis*.

87. A host cell according to claim 86 wherein the *Candida tropicalis* is *Candida tropicalis* 20336.

88. A host cell according to claim 87 wherein the *Candida tropicalis* is H5343 ura-.

89. A method of producing a *CYP52A3A* protein including an amino acid sequence as set forth in SEQ ID NO: 98 comprising:

a) transforming a suitable host cell with a DNA sequence that encodes the protein having the amino acid sequence as set forth in SEQ ID NO: 98; and

b) culturing the cell under conditions favoring the expression of the protein.

90. The method according to claim 89 wherein the step of culturing the cell comprises adding an organic substrate to media containing the cell.

91. Isolated nucleic acid encoding a *CYP52A3B* protein having the amino acid sequence as set forth in SEQ ID NO: 99.

92. Isolated nucleic acid comprising a coding region defined by nucleotides 913-2535 as set forth in SEQ ID NO: 89.

93. Isolated nucleic acid according to claim 92 comprising the nucleotide sequence as set forth in SEQ ID NO: 89.

94. Isolated protein comprising an amino acid sequence as set forth in SEQ ID NO: 99.

95. A vector comprising a nucleotide sequence encoding *CYP52A3B* protein including an amino acid sequence as set forth in SEQ ID NO: 99.

96. A vector according to claim 95 wherein the nucleotide sequence is set forth in nucleotides 913-2535 of SEQ ID NO: 89.

97. A vector according to claim 95 wherein the vector is selected from the group consisting of plasmid, phagemid, phage and cosmid.

98. A host cell transfected or transformed with the nucleic acid of claim 91.

99. A host cell according to claim 98 wherein the host cell is a yeast cell.

100. A host cell according to claim 99 wherein the yeast cell is a *Candida sp.*

101. A host cell according to claim 100 wherein the *Candida sp.* is *Candida tropicalis.*

102. A host cell according to claim 101 wherein the *Candida tropicalis* is *Candida tropicalis* 20336.

103. A host cell according to claim 102 wherein the *Candida tropicalis* is H5343 ura-.

104. A method of producing a *CYP52A3B* protein including an amino acid sequence as set forth in SEQ ID NO: 99 comprising:

a) transforming a suitable host cell with a DNA sequence that encodes the protein having the amino acid sequence as set forth in SEQ ID NO: 99; and

b) culturing the cell under conditions favoring the expression of the protein.

105. The method according to claim 104 wherein the step of culturing the cell comprises adding an organic substrate to media containing the cell.

106. Isolated nucleic acid encoding a *CYP52A5A* protein having the amino acid sequence set forth in SEQ ID NO: 100.

107. Isolated nucleic acid comprising a coding region defined by nucleotides 1103-2656 as set forth in SEQ ID NO: 90.

108. Isolated nucleic acid according to claim 107 comprising the nucleotide sequence as set forth in SEQ ID NO: 90.

109. Isolated protein comprising an amino acid sequence as set forth in SEQ ID NO: 100.

110. A vector comprising a nucleotide sequence encoding *CYP52A5A* protein including an amino acid sequence as set forth in SEQ ID NO: 100.

111. A vector according to claim 110 wherein the nucleotide sequence is set forth in nucleotides 1103-2656 OF SEQ ID NO: 90.

112. A vector according to claim 110 wherein the vector is selected from the group consisting of plasmid, phagemid, phage and cosmid.

113. A host cell transfected or transformed with the nucleic acid of claim 106.

114. A host cell according to claim 113 wherein the host cell is a yeast cell.

115. A host cell according to claim 114 wherein the yeast cell is a *Candida sp.*

116. A host cell according to claim 115 wherein the *Candida sp.* is *Candida tropicalis.*

117. A host cell according to claim 116 wherein the *Candida tropicalis* is *Candida tropicalis* 20336.

118. A host cell according to claim 117 wherein the *Candida tropicalis* is H5343 ura-.

119. A method of producing a *CYP52A5A* protein including an amino acid sequence as set forth in SEQ ID NO: 100 comprising:

a) transforming a suitable host cell with a DNA sequence that encodes the protein having the amino acid sequence as set forth in SEQ ID NO: 100; and

b) culturing the cell under conditions favoring the expression of the protein.

120. The method according to claim 119 wherein the step of culturing the cell comprises adding an organic substrate to media containing the cell.

121. Isolated nucleic acid encoding a *CYP52A5B* protein having the amino acid sequence as set forth in SEQ ID NO: 101.

122. Isolated nucleic acid comprising a coding region defined by nucleotides 1142-2695 as set forth in SEQ ID NO: 91.

123. Isolated nucleic acid according to claim 122 comprising the nucleotide sequence as set forth in SEQ ID NO: 91.

124. Isolated protein comprising an amino acid sequence as set forth in SEQ ID NO: 101.

125. A vector comprising a nucleotide sequence encoding *CYP52A5B* protein including the amino acid sequence as set forth in SEQ ID NO: 101.

126. A vector according to claim 125 wherein the nucleotide sequence is set forth in nucleotides 1142-2695 of SEQ ID NO: 91.

127. A vector according to claim 125 wherein the vector is selected from the group consisting of plasmid, phagemid, phage and cosmid.

128. A host cell transfected or transformed with the nucleic acid of claim 121.

129. A host cell according to claim 128 wherein the host cell is a yeast cell.

130. A host cell according to claim 129 wherein the yeast cell is a *Candida sp.*

131. A host cell according to claim 130 wherein the *Candida sp.* is *Candida tropicalis*.

132. A host cell according to claim 131 wherein the *Candida tropicalis* is *Candida tropicalis* 20336.

133. A host cell according to claim 132 wherein the *Candida tropicalis* is H5343 ura-.

134. A method of producing a *CYP52A5B* protein including an amino acid sequence as set forth in SEQ ID NO: 101 comprising:

a) transforming a suitable host cell with a DNA sequence that encodes the protein having the amino acid sequence as set forth in SEQ ID NO: 101; and

b) culturing the cell under conditions favoring the expression of the protein.

135. The method according to claim 134 wherein the step of culturing the cell comprises adding an organic substrate to media containing the cell.

136. Isolated nucleic acid encoding a *CYP52A8A* protein having the amino acid sequence set forth in SEQ ID NO: 102.

137. Isolated nucleic acid comprising a coding region defined by nucleotides 464-2002 as set forth in SEQ ID NO: 92.

138. Isolated nucleic acid according to claim 137 comprising the nucleotide sequence as set forth in SEQ ID NO: 92.

139. Isolated protein comprising an amino acid sequence as set forth in SEQ ID NO: 102.

140. A vector comprising a nucleotide sequence encoding *CYP52A8A* protein including an amino acid sequence as set forth in SEQ ID NO: 102.

141. A vector according to claim 140 wherein the nucleotide sequence is set forth in nucleotides 464-2002 of SEQ ID NO: 92.

142. A vector according to claim 140 wherein the vector is selected from the group consisting of plasmid, phagemid, phage and cosmid.

143. A host cell transfected or transformed with the nucleic acid of claim 136.

144. A host cell according to claim 143 wherein the host cell is a yeast cell.

145. A host cell according to claim 144 wherein the yeast cell is a *Candida sp.*

146. A host cell according to claim 145 wherein the *Candida sp.* is *Candida tropicalis.*

147. A host cell according to claim 146 wherein the *Candida tropicalis* is *Candida tropicalis* 20336.

148. A host cell according to claim 147 wherein the *Candida tropicalis* is H5343 ura-.

149. A method of producing a *CYP52A8A* protein including an amino acid sequence as set forth in SEQ ID NO: 102 comprising:

a) transforming a suitable host cell with a DNA sequence that encodes the protein having the amino acid sequence as set forth in SEQ ID NO: 102; and

b) culturing the cell under conditions favoring the expression of the protein.

150. The method according to claim 149 wherein the step of culturing the cell comprises adding an organic substrate to media containing the cell.

151. Isolated nucleic acid encoding a *CYP52A8B* protein having the amino acid sequence set forth in SEQ ID NO: 103.

152. Isolated nucleic acid comprising a coding region defined by nucleotides 1017-2555 as set forth in SEQ ID NO: 93.

153. Isolated nucleic acid according to claim 152 comprising the nucleotide sequence as set forth in SEQ ID NO: 93.

154. Isolated protein comprising an amino acid sequence as set forth in SEQ ID NO: 103.

155. A vector comprising a nucleotide sequence encoding *CYP52A8B* protein including an amino acid sequence as set forth in SEQ ID NO: 103.

156. A vector according to claim 155 wherein the nucleotide sequence is set forth in nucleotides 1017-2555 of SEQ ID NO: 93.

157. A vector according to claim 155 wherein the vector is selected from the group consisting of plasmid, phagemid, phage and cosmid.

158. A host cell transfected or transformed with the nucleic acid of claim 151.

159. A host cell according to claim 158 wherein the host cell is a yeast cell.

160. A host cell according to claim 159 wherein the yeast cell is a *Candida sp.*

161. A host cell according to claim 160 wherein the *Candida sp.* is *Candida tropicalis.*

162. A host cell according to claim 161 wherein the *Candida tropicalis* is *Candida tropicalis* 20336.

163. A host cell according to claim 162 wherein the *Candida tropicalis* is H5343 ura-.

164. A method of producing a *CYP52A8B* protein including an amino acid sequence as set forth in SEQ ID NO: 103 comprising:

a) transforming a suitable host cell with a DNA sequence that encodes the protein having the amino acid sequence as set forth in SEQ ID NO: 103; and

b) culturing the cell under conditions favoring the expression of the protein.

165. The method according to claim 164 wherein the step of culturing the cell comprises adding an organic substrate to media containing the cell.

166. Isolated nucleic acid encoding a *CYP52D4A* protein having the amino acid sequence set forth in SEQ ID NO: 104.

167. Isolated nucleic acid comprising a coding region defined by nucleotides 767-2266 as set forth in SEQ ID NO: 94.

168. Isolated nucleic acid according to claim 167 comprising the nucleotide sequence as set forth in SEQ ID NO: 94.

169. Isolated protein comprising an amino acid sequence as set forth in SEQ ID NO: 104.

170. A vector comprising a nucleotide sequence encoding *CYP52D4A* protein including an amino acid sequence as set forth in SEQ ID NO: 104.

171. A vector according to claim 170 wherein the nucleotide sequence is set forth in nucleotides 767-2266 of SEQ ID NO: 94.

172. A vector according to claim 170 wherein the vector is selected from the group consisting of plasmid, phagemid, phage and cosmid.

173. A host cell transfected or transformed with the nucleic acid of claim 166.

174. A host cell according to claim 173 wherein the host cell is a yeast cell.

175. A host cell according to claim 174 wherein the yeast cell is a *Candida sp.*

176. A host cell according to claim 175 wherein the *Candida sp.* is *Candida tropicalis.*

177. A host cell according to claim 176 wherein the *Candida tropicalis* is *Candida tropicalis* 20336.

178. A host cell according to claim 177 wherein the *Candida tropicalis* is H5343 ura-.

179. A method of producing a *CYP52D4A* protein including an amino acid sequence as set forth in SEQ ID NO: 104 comprising:

a) transforming a suitable host cell with a DNA sequence that encodes the protein having the amino acid sequence as set forth in SEQ ID NO: 104; and

b) culturing the cell under conditions favoring the expression of the protein.

180. The method according to claim 179 wherein the step of culturing the cell comprises adding an organic substrate to media containing the cell.

181. A method for discriminating members of a gene family by quantifying the amount of target mRNA in a sample comprising:

a) providing an organism containing a target gene;

b) culturing the organism with an organic substrate which causes upregulation in the activity of the target gene;

c) obtaining a sample of total RNA from the organism at a first point in time;

d) combining at least a portion of the sample of the total RNA with a known amount of competitor RNA to form an RNA mixture, wherein the competitor RNA is substantially similar to the target mRNA but has a lesser number of nucleotides compared to the target mRNA;

e) adding reverse transcriptase to the RNA mixture in a quantity sufficient to form corresponding target DNA and competitor DNA;

f) conducting a polymerase chain reaction in the presence of at least one primer specific for at least one substantially non-homologous region of the target DNA within the gene family, the primer also specific for the competitor DNA;

g) repeating steps (c-f) using increasing amounts of the competitor RNA while maintaining a substantially constant amount of target RNA;

(h) determining the point at which the amount of target DNA is substantially equal to the amount of competitor DNA;

(i) quantifying the results by comparing the ratio of the concentration of unknown target to the known concentration of competitor; and

(j) obtaining a sample of total RNA from the organism at another point in time and repeating steps (d-i).

182. A method according to claim 181 wherein the target gene is selected from the group consisting of a *CPR* gene and a *CYP* gene.

183. A method according to claim 182 wherein the *CPR* gene is selected from the group consisting of a *CPRA* gene (SEQ ID NO: 81) and a *CPRB* gene (SEQ ID NO: 82).

184. A method according to claim 182 wherein the *CYP* gene is selected from the group consisting of *CYP52A1A* gene (SEQ ID NO: 85), *CYP52A2A* gene (SEQ ID NO: 86), *CYP52A2B* gene (SEQ ID NO: 87), *CYP52A3A* gene (SEQ ID NO: 88), *CYP52A3B* gene (SEQ ID NO. 89), *CYP52A5A* gene (SEQ ID NO: 90), *CYP52A5B* gene (SEQ ID NO: 91), *CYP52A8A* gene (SEQ ID NO: 92), *CYP52A8B* gene (SEQ ID NO: 93) and *CYP52D4A* gene (SEQ ID NO: 94).

185. A method for increasing production of a dicarboxylic acid comprising:

a) providing a host cell having a naturally occurring number of *CPRA* genes;

b) increasing, in the host cell, the number of *CPRA* genes which encode a *CPRA* protein having the amino acid sequence as set forth in SEQ ID NO: 83;

c) culturing the host cell in media containing an organic substrate which upregulates the *CPRA* gene, to effect increased production of dicarboxylic acid.

186. A method for increasing the production of a *CPRA* protein having an amino acid sequence as set forth in SEQ ID NO: 83 comprising:

a) transforming a host cell having a naturally occurring amount of *CPRA* protein with an increased copy number of a *CPRA* gene that encodes the *CPRA* protein having the amino acid sequence as set forth in SEQ ID NO: 83; and

b) culturing the cell and thereby increasing expression of the protein compared with that of a host cell containing a naturally occurring copy number of the *CPRA* gene.

187. A method for increasing production of a dicarboxylic acid comprising:

a) providing a host cell having a naturally occurring number of *CPRB* genes;

b) increasing, in the host cell, the number of *CPRB* genes which encode a *CPRB* protein having the amino acid sequence as set forth in SEQ ID NO: 84;

c) culturing the host cell in media containing an organic substrate which upregulates the *CPRB* gene, to effect increased production of dicarboxylic acid.

188. A method for increasing the production of a *CPRB* protein having an amino acid sequence as set forth in SEQ ID NO: 84 comprising:

a) transforming a host cell having a naturally occurring amount of *CPRB* protein with an increased copy number of a *CPRB* gene that encodes the *CPRB* protein having the amino acid sequence as set forth in SEQ ID NO: 84; and

b) culturing the cell and thereby increasing expression of the protein compared with that of a host cell containing a naturally occurring copy number of the *CPRB* gene.

189. A method for increasing production of a dicarboxylic acid comprising:

a) providing a host cell having a naturally occurring number of *CYP52A1A* genes;

b) increasing, in the host cell, the number of *CYP52A1A* genes which encode a *CYP52A1A* protein having the amino acid sequence as set forth in SEQ ID NO: 95;

c) culturing the host cell in media containing an organic substrate which upregulates the *CYP52A2A* gene, to effect increased production of dicarboxylic acid.

190. A method for increasing the production of a *CYP52A1A* protein having an amino acid sequence as set forth in SEQ ID NO: 95 comprising:

a) transforming a host cell having a naturally occurring amount of *CYP52A1A* protein with an increased copy number of a *CYP52A1A* gene that encodes the *CYP52A1A* protein having the amino acid sequence as set forth in SEQ ID NO: 95; and

b) culturing the cell and thereby increasing expression of the protein compared with that of a host cell containing a naturally occurring copy number of the *CYP52A1A* gene.

191. A method for increasing production of a dicarboxylic acid comprising:

a) providing a host cell having a naturally occurring number of *CYP52A2A* genes;

b) increasing, in the host cell, the number of *CYP52A2A* genes which encode a *CYP52A2A* protein having the amino acid sequence as set forth in SEQ ID NO: 96;

c) culturing the host cell in media containing an organic substrate which upregulates the *CYP52A2A* gene, to effect increased production of dicarboxylic acid.

192. A method for increasing the production of a *CYP52A2A* protein having an amino acid sequence as set forth in SEQ ID NO: 96 comprising:

a) transforming a host cell having a naturally occurring amount of *CYP52A2A* protein with an increased copy number of a *CYP52A2A* gene that encodes the *CYP52A2A* protein having the amino acid sequence as set forth in SEQ ID NO: 96; and

b) culturing the cell and thereby increasing expression of the protein compared with that of a host cell containing a naturally occurring copy number of the *CYP52A2A* gene.

193. A method for increasing production of a dicarboxylic acid comprising:

a) providing a host cell having a naturally occurring number of *CYP52A2B* genes;

b) increasing, in the host cell, the number of *CYP52A2B* genes which encode a *CYP52A2B* protein having the amino acid sequence as set forth in SEQ ID NO: 97;

c) culturing the host cell in media containing an organic substrate which upregulates the *CYP52A2B* gene, to effect increased production of dicarboxylic acid.

194. A method for increasing the production of a *CYP52A2B* protein having an amino acid sequence as set forth in SEQ ID NO: 97 comprising:

a) transforming a host cell having a naturally occurring amount of *CYP52A2B* protein with an increased copy number of a *CYP52A2B* gene that encodes the *CYP52A2B* protein having the amino acid sequence as set forth in SEQ ID NO: 97; and

b) culturing the cell and thereby increasing expression of the protein compared with that of a host cell containing a naturally occurring copy number of the *CYP52A2B* gene.

195. A method for increasing production of a dicarboxylic acid comprising:

a) providing a host cell having a naturally occurring number of *CYP52A3A* genes;

b) increasing, in the host cell, the number of *CYP52A3A* genes which encode a *CYP52A3A* protein having the amino acid sequence as set forth in SEQ ID NO: 98;

c) culturing the host cell in media containing an organic substrate which upregulates the *CYP52A3A* gene, to effect increased production of dicarboxylic acid.

196. A method for increasing the production of a *CYP52A3A* protein having an amino acid sequence as set forth in SEQ ID NO: 98 comprising:

a) transforming a host cell having a naturally occurring amount of *CYP52A3A* protein with an increased copy number of a *CYP52A3A* gene that encodes the *CYP52A3A* protein having the amino acid sequence as set forth in SEQ ID NO: 98; and

b) culturing the cell and thereby increasing expression of the protein compared with that of a host cell containing a naturally occurring copy number of the *CYP52A3A* gene.

197. A method for increasing production of a dicarboxylic acid comprising:

a) providing a host cell having a naturally occurring number of *CYP52A3B* genes;

b) increasing, in the host cell, the number of *CYP52A3B* genes which encode a *CYP52A3B* protein having the amino acid sequence as set forth in SEQ ID NO: 99;

c) culturing the host cell in media containing an organic substrate which upregulates the *CYP52A3B* gene, to effect increased production of dicarboxylic acid.

198. A method for increasing the production of a *CYP52A3B* protein having an amino acid sequence as set forth in SEQ ID NO: 99 comprising:

a) transforming a host cell having a naturally occurring amount of *CYP52A3B* protein with an increased copy number of a *CYP52A3B* gene that encodes the *CYP52A3B* protein having the amino acid sequence as set forth in SEQ ID NO: 99; and

b) culturing the cell and thereby increasing expression of the protein compared with that of a host cell containing a naturally occurring copy number of the *CYP52A3B* gene.

199. A method for increasing production of a dicarboxylic acid comprising:

a) providing a host cell having a naturally occurring number of *CYP52A5A* genes;

b) increasing, in the host cell, the number of *CYP52A5A* genes which encode a *CYP52A5A* protein having the amino acid sequence as set forth in SEQ ID NO: 100;

c) culturing the host cell in media containing an organic substrate which upregulates the *CYP52A5A* gene, to effect increased production of dicarboxylic acid.

200. A method for increasing the production of a *CYP52A5A* protein having an amino acid sequence as set forth in SEQ ID NO: 100 comprising:

a) transforming a host cell having a naturally occurring amount of *CYP52A5A* protein with an increased copy number of a *CYP52A5A* gene that encodes the *CYP52A5A* protein having the amino acid sequence as set forth in SEQ ID NO: 100; and

b) culturing the cell and thereby increasing expression of the protein compared with that of a host cell containing a naturally occurring copy number of the *CYP52A5A* gene.

201. A method for increasing production of a dicarboxylic acid comprising:

a) providing a host cell having a naturally occurring number of *CYP52A5B* genes;

b) increasing, in the host cell, the number of *CYP52A5B* genes which encode a *CYP52A5B* protein having the amino acid sequence as set forth in SEQ ID NO: 101;

c) culturing the host cell in media containing an organic substrate which upregulates the *CYP52A5B* gene, to effect increased production of dicarboxylic acid.

202. A method for increasing the production of a *CYP52A5B* protein having an amino acid sequence as set forth in SEQ ID NO: 101 comprising:

a) transforming a host cell having a naturally occurring amount f *CYP52A5B* protein with an increased copy number of a *CYP52A5B* gene that encodes the *CYP52A5B* protein having the amino acid sequence as set forth in SEQ ID NO: 101; and

b) culturing the cell and thereby increasing expression of the protein compared with that of a host cell containing a naturally occurring copy number of the *CYP52A5B* gene.

203. A method for increasing production of a dicarboxylic acid comprising:

a) providing a host cell having a naturally occurring number of *CYP52A8A* genes;

b) increasing, in the host cell, the number of *CYP52A8A* genes which encode a *CYP52A8A* protein having the amino acid sequence as set forth in SEQ ID NO: 102;

c) culturing the host cell in media containing an organic substrate which upregulates the *CYP52A8A* gene, to effect increased production of dicarboxylic acid.

204. A method for increasing the production of a *CYP52A8A* protein having an amino acid sequence as set forth in SEQ ID NO: 102 comprising:

a) transforming a host cell having a naturally occurring amount of *CYP52A8A* protein with an increased copy number of a *CYP52A8A* gene that encodes the *CYP52A8A* protein having the amino acid sequence as set forth in SEQ ID NO: 102; and

b) culturing the cell and thereby increasing expression of the protein compared with that of a host cell containing a naturally occurring copy number of the *CYP52A8A* gene.

205. A method for increasing production of a dicarboxylic acid comprising:

a) providing a host cell having a naturally occurring number of *CYP52A8B* genes;

b) increasing, in the host cell, the number of *CYP52A8B* genes which encode a *CYP52A8B* protein having the amino acid sequence as set forth in SEQ ID NO: 103;

c) culturing the host cell in media containing an organic substrate which upregulates the *CYP52A8B* gene, to effect increased production of dicarboxylic acid.

206. A method for increasing the production of a *CYP52A8B* protein having an amino acid sequence as set forth in SEQ ID NO: 103 comprising:

a) transforming a host cell having a naturally occurring amount of *CYP52A8B* protein with an increased copy number of a *CYP52A8B* gene that encodes the *CYP52A8B* protein having the amino acid sequence as set forth in SEQ ID NO: 103; and

b) culturing the cell and thereby increasing expression of the protein compared with that of a host cell containing a naturally occurring copy number of the *CYP52A8B* gene.

207. A method for increasing production of a dicarboxylic acid comprising:

a) providing a host cell having a naturally occurring number of *CYP52D4A* genes;

b) increasing, in the host cell, the number of *CYP52D4A* genes which encode a *CYP52D4A* protein having the amino acid sequence as set forth in SEQ ID NO: 104;

c) culturing the host cell in media containing an organic substrate which upregulates the *CYP52D4A* gene, to effect increased production of dicarboxylic acid.

208. A method for increasing the production of a *CYP52D4A* protein having an amino acid sequence as set forth in SEQ ID NO: 104 comprising:

a) transforming a host cell having a naturally occurring amount of *CYP52D4A* protein with an increased copy number of a *CYP52D4A* gene that encodes the *CYP52D4A* protein having the amino acid sequence as set forth in SEQ ID NO: 104; and

b) culturing the cell and thereby increasing expression of the protein compared with that of a host cell containing a naturally occurring copy number of the *CYP52D4A* gene.

209. A method for discriminating members of a gene family according to claim 181 wherein culturing the organism with the organic substrate is accomplished in a fermentor.

# ABSTRACT

Novel genes have been isolated which encode cytochrome P450 and NADPH reductase enzymes of the ω-hydroxylase complex of *C. tropicalis* 20336. Vectors including these genes, transfected host cells and transformed host cells are provided. Methods of producing of cytochrome P450 and NADPH reductase enzymes are also provided which involve transforming a host cell with a gene encoding these enzymes and culturing the cells. Methods of increasing the production of a dicarboxylic acid and methods of increasing production of the aforementioned enzymes are also provided which involve increasing in the host cell the number of genes encoding these enzymes. A method for discriminating members of a gene family by quantifying the expression of genes is also provided.